Poellnitz, Gerhard von

Graphische darstellung und ihre anwendung auf...

München

1914

COLUMBIA UNIVERSITY LIBRARIES
PRESERVATION DIVISION

BIBLIOGRAPHIC MICROFORM TARGET

ORIGINAL MATERIAL AS FILMED - EXISTING BIBLIOGRAPHIC RECORD

3
Box 48 Poellnitz, Gerhard von, 1876-
Graphische darstellung und ihre anwendung auf die statistik der veranlagten einkommen. München, Honig, 1914.
76 p. diagrs., tables., 26 cm.

Thesis, Münich.

TECHNICAL MICROFORM DATA

FILM SIZE: 35mm REDUCTION RATIO: 14:1 IMAGE PLACEMENT: IA (IIA) IB IIB

DATE FILMED: 10-16-97 INITIALS: AP

TRACKING # : 28439

FILMED BY PRESERVATION RESOURCES, BETHLEHEM, PA.

Graphische Darstellung und ihre Anwendung auf die Statistik der veranlagten Einkommen.

Inaugural-Dissertation

der

Hohen staatswirtschaftlichen Fakultät

der

Königl. Ludwig-Maximilians-Universität

zur

Erlangung der Doktorwürde.

Vorgelegt von

Dr. G. v. Poellnitz.

MÜNCHEN 1914.
Ernst Honig, Buchhandlung und Verlag
Frauenstr. 11

Angenommen von der staatswirtschaftlichen Fakultät.

Referenten: Unterstaatssekretär z. D. Prof. Dr. G. v. Mayr
und Geh. Hofrat Prof. Dr. F. Lindemann.

Graphische Darstellung und ihre Anwendung auf die Statistik der veranlagten Einkommen.

Inaugural-Dissertation

der

Hohen staatswirtschaftlichen Fakultät

der

Königl. Ludwig-Maximilians-Universität

zur

Erlangung der Doktorwürde.

vorgelegt von

Dr. G. v. Poellnitz.

MÜNCHEN 1914.
Ernst Honig, Buchhandlung und Verlag
[illegible]

Angenommen von der staatswirtschaftlichen Fakultät.
Referenten: Unterstaatssekretär z. D. Prof. Dr. G. v. Mayr
und Geh. Hofrat Prof. Dr. F. Lindemann.

Graphische Darstellung und ihre Anwendung auf die Statistik der veranlagten Einkommen.

Inaugural-Dissertation

der

Hohen staatswirtschaftlichen Fakultät

der

Königl. Ludwig-Maximilians-Universität

zur

Erlangung der Doktorwürde.

Vorgelegt von

Dr. G. v. Poellnitz.

MÜNCHEN 1914.
Ernst Honig, Buchhandlung und Verlag
Frauenstr. 11

Inhaltsübersicht

Literaturverzeichnis.

Otto Ammon, „Die Gesellschaftsordnung und ihre natürlichen Grundlagen". 3. Aufl. Jena 1900.

L. Amoroso, La distribuzione della ricchezza come fenomeno die distribuzione Giornale degli Ecnomisti, IX. 12. Roma 1912.

Felix Auerbach, Die graphische Darstellung, Leipzig 1914. (Aus „Natur und Geisteswelt", 437.)

Rodolfo Benini, „Di alcune curve descritte da fenomeni economici in Giornale degli economisti 1897.

— Prinipi di demografia. Firenze 1901.

— I diagrammi a scala logaritmica in den „Festgaben für Adolf Wagner". S. 119. Leipzig 1905.

— Principi di Statistica metodologica. Torino 1906. S. 146. Diagrama a doppia scala logaritmica.

Bertillon, Bericht im Bulletin de l'institut international de Statistique. Session Budapest XIII. S. 313.

Ladislaus Bortkiewicz, „Die Grenznutzentheorie als Grundlage einer ultraliberalen Wirtschaftspolitik" in Schmollers Jahrbuch 1898, IV. Heft, S. 89.

C. Bresciani-Turroni, Alcuni appunti sulla distribuzione del reddito e del patrimonoi in Prussia. In „Festgaben zum 70. Geburtstag Adolf Wagners". S. 185. Leipzig 1905. (Geht von der Trennung von Besitz und Arbeitseinkommen aus.)

— Über Methoden der Einkommensverteilungsstatistik . Jahrbuch für National-Ökonomie und Statistik, 1907.

Moritz Cantor, Politische Arithmetik oder die Arithmetik des täglichen Lebens. 2. Aufl. Leipzig 1903.

Edgeworth, im Journal of the Royal Statistical Society. September 1896.)

P. des Essars, Vortrag. Journal de la Société de statistique de Paris 1902.

Gustav Theodor Fechner, „Kollektivmaßlehre" (herausgegeben von Lipps). Leipzig 1897.

V. Furlan, Note sulla curva paretiana dei redditi Giornale degli Economisti, Roma 1909.

— Neuere Literatur zur Einkommensverteilung in Italien: Corrado Gini in Conrads Jahrbücher, August 1911.

Antonio Gabaglio, Teoria generale della statistica Milano 1888.

Corrado Gini, Il diverso accrescimento delle classe sociali e la concentrazione della ricchezza. Giornale degli Economisti Roma, gennaio 1909.

— Indici di concentrazione e di dipendenza in Atti della Societa Italiana per il progressa delle Scienze 3 S. 453. Padova 1909.

Hasse, Herm., Die Statistik als Hilfswissenschaft der Sozialwissenschaften. Gautzsch 1911.

Hochsinger, Frau Dr. phil., Referat über das sogen. Pareto'sche Gesetz in Statistische Monatsschrift N F XVI. S. 632 ff. Brünn 1911.

Al. Kaufmann, Theorie und Methoden der Statistik . Petersburg 1913.

Kiaer, Repartition sociale des revenus, Bull. de l'Inst. int. de St. 1909.

A. Liesse, La statistique. Paris 1905.

Mayet, Session de Berlin. Citiert bei L. March, Rapport à l'institut 1911, XIX.

Lucien March, quelques exemples de distribution de salaires, Journal de la Soc. de St. de Paris. VI. VII. 1898.

v. Mayr, „Theoretische Statistik". Freiburg 1895.

R. Meyer, Art.: „Einkommen", II. Statistik des Einkommens und der Einkommens-Verteilung im Handwörterbuch der Staatswissenschaften. III. Aufl. III. Bd. S. 663.

R. Neuendorf, Praktische Mathematik. I. Graphisches und nummerisches Rechnen. Leipzig 1911. (Aus „Natur und Geisteswelt".)

V. Pareto, in Giornale degli Econositi
1. La legge della domanda. 1895.
2. Sul modo di figurare i fenomeni economici. 1896.
3. La curva delle entrate e le osservazioni del prof. Edgeworth, November 1896.
4. Aggiunta allo studio sulla curva delle entrate. 1897.

V. Pareto, Cours d'économie politique Bd. 2. Lausanne. 1897. Kapitel „La courbe des revenus. §§ 950 ff.

— Nouvelle methode d'interpolation pour les phenomènes données par l'experience; in „Rapports et comptes rendus du VIme Congrès international de psychologie. Génève 1910.

Dr. med. Roesle, Vortrag in der deutschen Statistischen Gesellschaft in der zweiten Mitgliederversammlung am 22.—23. Oktober 1912 in Berlin.

S. Schott, Die Statistik. Leipzig 1905.

— „Graphische Darstellungen" in dem Werk „Die Statistik in Deutschland". Band I. München 1912.

Tabacovici, Methoden der Einkommensteuerstatistik. Leipzig 1913.

Giuseppe Tammeo, La statistica. Torino 1896.

Dr. Emanuel Hugo Vogel, Zur Methodologie der Einkommenstatistik. (Österr.) Statistische Monatsschrift, XVII, S. 495. Brünn 1912.

Dr. Franz Žižek, Die statistischen Mittelwerte. Leipzig 1908.



A. Graphische Darstellung.

1. Notwendigkeit zusammenfassender Bearbeitung der statistischen Resultate

Das Anwachsen des statistichen Urmaterials aus bescheidenen Anfängen zu seinem heutigen sachlichen, räumlichen und zeitlichen Umfang macht eine zusammenfassende und vergleichende Bearbeitung immer unentbehrlicher.

Diese wird sich naturgemäß zunächst nach dem Wesen des erforschten Gegenstandes zu richten haben, sie wird die Bedeutung der statistisch festgestellten Tatsachen würdigen, wird untersuchen, ob sich aus den Ergebnissen der Statistik Schlüsse auf einen kausalen Zusammenhang der Ereignisse und Zustände ziehen lassen, oder ob sie einen aus anderen Gründen vermuteten Kausalzusammenhang bestätigen.

In gewisser Hinsicht haben alle Arten der Bearbeitung gemeinsame Züge. Die Statistik als Massenbeobachtung bringt ihre Ergebnisse in die Form von Zahlen und Zahlenbeziehungen. Für die Inbeziehungsetzung von Zahlen hat die Mathematik eine Methode ausgebildet, die von den einfachen Rechnungsarten ausgehend zu immer größerer Kompliziertheit ansteigt. Diese Methode ist anwendbar, soweit es sich nur um die Zahlenbeziehungen als solche handelt.

2. Graphische Darstellung anwendbar, wenn die Vorteile die Nachteile überwiegen

Geht man einen Schritt weiter und drückt die zahlenmäßigen Ergebnisse der Statistik graphisch durch räumliche Größen aus, so wird eine neue Denkoperation vom Leser und Beschauer verlangt, um dann andere Denkoperationen zu ersparen und zu Ergebnissen zu gelangen, die auf anderem Wege nur schwer oder gar nicht zu erreichen sind. Die geistige Ökonomie verlangt, daß der Aufwand von neuen Erschwerungen so gering als möglich bemessen wird, und daß er unterbleibt, wenn die Vorteile ihn nicht überwiegen.

Die neue Schwierigkeit besteht zunächst darin, daß die Zahlen von Jugend auf gelernt und im täglichen Gebrauch geübt worden sind, daß sich mit ihnen unwillkürlich Vorstellungen verbinden. Die Anschauung von räumlichen Größen ist nicht so allgemein entwickelt und es besteht keine so unmittelbare, unwillkürliche Ideenassoziation zwischen der räumlichen Größe und dem Begriff der Zahl, als bei dem Anblick der gewohnten Ziffern. Damit nun diese Ideenassoziation nicht verloren geht, empfiehlt es sich, den gewählten Maßstab nicht nur durch Darstellung der Einheit, sondern auch durch eine Skala am Rande zu bezeichnen, und so weit als möglich auf der Figur die dargestellten Zahlen anzubringen.[1]) So ist die unmittelbare Verständlichkeit der Tabelle mit der Übersichtlichkeit und Anschaulichkeit des Diagrammes vereinigt.

Ein weiterer Mangel der graphischen Darstellungen scheint zu sein, daß sich auf sie die Operationen der Arithmetik und Algebra nicht in der Weise anwenden lassen, wie auf Ziffern. Zum Teil liegt hierfür ein Ersatz in der Möglichkeit, durch geometrische Konstruktion dieselbe Wirkung zu erreichen, z. B. Addition und Subtraktion durch Anfügen und Abtrennen von Strecken und Flächen, Multiplikation durch Konstruktion eines Rechtecks aus seinen Seiten, Addition und Substraktion von Quadraten und andern unter sich ähn-

[1]) Roesle. Vortrag a. a. O.

lichen Figuren mit Hilfe des pythagroräischen Lehrsatzes usw. Zuweilen ist die Konstruktion bequemer als die Rechnung; aber auch wo die Rechnung zweckmäßiger erscheint, bleibt es immer möglich, den Endergebnissen der Rechnung wieder die graphische Form zu geben. Ein prinzipielles Hindernis steht dem nicht entgegen.

Allerdings ist die Verständlichkeit größer, wenn die ursprünglichen Zahlen graphisch wiedergegeben werden. Denn sie messen Gegenstände oder Vorgänge, die so deutlich in Begriffe zu fassen und zu erkennen sind, daß sie einer Zählung unterworfen werden konnten. Bei den abgeleiteten Zahlen trifft dies nicht immer zu. Man kann sich zwar einen prozentualen Zuwachs konkret als die Menge vorstellen, die zu 100 Einheiten hinzukommt, bei Logarithmen, Potenzen und Wurzeln und bei zusammengesetzten Rechnungsformen ist dagegen eine konkrete Vorstellung in der Regel nicht möglich. Andererseits ist bereits in der Zahl als solcher eine Abstraktion enthalten und dem Geist geläufig; man ist gewohnt, nicht nur mit der Zahl als Menge bestimmter Objekte, sondern als Größe an sich zu arbeiten, auch wenn die Einheiten, die sie zählt, nicht vorstellbar sind. Die Zahlen an sich sind es, und diese können stets durch die Zeichnung wiedergegeben werden.

So ist theoretisch jede beliebige Ableitung aus einzelnen oder mehreren der beobachteten Zahlen graphisch darstellbar, sei es durch Konstruktion oder durch bloße Darstellung des Resultates. Die Entscheidung, ob Konstruktion oder Darstellung bloßer Rechnungsresultate vorzuziehen ist, wird mit Rücksicht auf den Zweck der Anschaulichkeit zu fällen sein. Es ist zu erwarten, daß diesem Zweck am besten eine Beschränkung auf die einfachsten geometrischen Formen entsprechen wird, die dem Auge eine unmittelbare Schätzung gestatten, und daß komplizierte geometrische Gebilde nur da am Platze sind, wo sie die Rechnung räumlich wiederspiegeln, und ihre Entstehung verständlich gemacht werden kann.

Auch diese Frage läßt sich nur lösen, wenn man die mannigfaltigen Formen, die der graphischen Darstellung zur Verfügung stehen, im Zusammenhang betrachtet, und zwar sowohl als Mittel, eine Größe zu veranschaulichen, als um sie zu vergleichen, sowohl nach ihrer geometrischen Konstruktion, als nach ihrer Beziehung zum dargestellten Objekt.

3. Elemente graphischer Darstellung

Betrachten wir zunächst die geometrische Konstruktion, so finden wir als einfachstes Gebilde den Punkt, sichtbar gemacht durch eine kleine schwarze oder farbige Fläche, die nicht durch ihre Ausdehnung, sondern nur durch ihre Existenz von Bedeutung ist. Sie ist das natürliche Symbol der Einheit, mehrere Punkte das Symbol ihrer Zahl. Punkte und ebenso andere Zeichen, die nicht gemessen, sondern gezählt werden sollen, können durch ihre Anordnung auf dem Papier die räumliche Verteilung der symbolisierten Ereignisse wiedergeben, z. B. die Todesfälle infolge einer Krankheit auf einem Stadtplan, Truppenkörper auf einer Karte usw. Wird von der geographischen Lage abgesehen, so werden die Punkte zweckmäßig so angeordnet, daß sie regelmäßige Figuren bilden, z. B. in gleichen Abständen auf einer geraden Linie stehen.

Damit kommen wir auf die gerade Linie, deren Länge der dargestellten Größe entspricht. Eine einzige gerade Linie ist außerstande, eine Größe darzustellen; zum mindesten muß die der Einheit entsprechende Länge zum Vergleich gegeben werden. In gleicher Weise, wie die erste dargestellte Größe mit der Einheit durch Längen von Linien verglichen wird, wird die erste mit jeder nachfolgenden Größe durch Längen verglichen. Diese Längen können auf derselben Linie abgemessen und bezeichnet werden; es können aber auch Linien verschiedener Länge, die den Zahlen proportional sind, nebeneinander gestellt werden. Bei letzterer Anordnung ergibt sich von selbst die Kombination zweier Richtungen von Linien. Auch wenn man die Grundlinie, auf der die Linien stehen, nicht zeichnet, kann sie leicht vorgestellt werden. Das Nächstliegende wird sein, die Linien senkrecht auf eine Gerade in gleichen Abständen aufzustellen. Diese gleichen Abstände können dann ebenfalls Größen ausdrücken. Ein häufiger Fall, der bereits zu einer feststehenden Konvention geworden ist, setzt die Abstände auf der Grundlinie als Symbole der Zeit, die zwischen den Beobachtungen derselben Erscheinung verlaufen ist. Nimmt man dies an, so muß die Proportion genau durchgeführt werden, und bei unregelmäßigen Abständen der Beobachtungszeit müssen verschiedene Abstände der Linien genommen werden, die alle unter sich und zu der Zeiteinheit im richtigen Verhältnis stehen. Der Abstand der Linien kann auch anderen Größen proportional sein. In jedem Fall erhält man eine Figur, die in der Mathematik als Koordinatensystem bezeichnet wird. Die Grundlinie ist die Abzissenachse; auf dieser wird ein Nullpunkt angenommen und nach rechts Längenabschnitte abgemessen, die Abszissen, auf deren Endpunkten die senkrechten Linien-Ordinaten errichtet werden. Bedeutet z. B. der auf der Abszissenachse gewählte Nullpunkt den Beginn eines Jahres, so sind die Beginne der folgenden Jahre in gleichen Abständen anzubringen. Die auf dem Nullpunkt senkrecht errichtete Linie ist die Ordinatenachse. Am zweckmäßigsten ist das echtwinkelige, cartesianische Koordinatensystem.

Vergleicht man die Längen der Ordinaten, so bewegt sich der Blick von einem Endpunkt zum nächsten auf einer vorgestellten Linie. Die Vergleichung wird erleichtert, wenn man diese Linie zeichnet. Man nennt diese Verbindungslinien in der Statistik häufig „Kurven", auch wenn sie in Wirklichkeit gebrochene, gerade Linien sind.[1]) In derselben Figur können mehrere Kurven eingezeichnet werden, und es ergeben sich dabei mannigfaltige Kombinationen, auf die später näher eingegangen werden soll.

Außer dem bekannten rechtwinkeligen Koordinatensystem sind auch solche in anderen Formen möglich, die alle die Abhängigkeit zweier veränderlicher Größen von einander zum Ausdruck bringen. Zum Beispiel kann statt des rechten Winkels ein spitzer oder stumpfer gewählt werden, und die Abszissen können durch Winkel oder Kreisbogen ersetzt werden. Letztere Form ist dann angemessen, wenn das erste und letzte Glied der Reihe zusammenfallen, z. B. bei Durchschnittszahlen für die Monate einer Reihe von Jahren. Die anschaulichste und leichtest verständliche Form ist jedoch das rechtwinklige Koordinatensystem.

Mehrere gerade Linien in einer Ebene können eine Fläche vollständig begrenzen, und damit wieder eine Größe zum Ausdruck bringen. Der Inhalt einer Fläche läßt sich durch die Winkel und Längen seiner Begrenzung bestimmen. Am einfachsten ist die Berechnung bei einem Rechteck. Die Flächen von Rechtecken verhalten sich, wie die Produkte zweier aufeinanderfolgender Seiten oder die Produkte von Grundlinie und Höhe. Mit Hilfe dieses Satzes kann man leicht das Produkt zweier Zahlen und diese selbst auf derselben

[1]) Die Verbindung der Punkte im Koordinatensystem ist nicht immer zweckmäßig; dann sind die Ordinaten unverbunden zu zeichnen oder nur die Punkte einzutragen. Siehe Näheres darüber S. 26 ff. und Anhang 4.

Figur graphisch darstellen, ebenso auch Dividend, Divisor und Quotient. So sind z. B. auf der Figur im 22. Jahrgang des Statistischen Jahrbuchs für das Deutsche Reich 1901 die Grundlinien der Rechtecke proportional der Erntefläche, die Höhe dem Ertrag pro ha und die Fläche proportional dem Gesamtbetrag der Ernte. Übrigens sind nicht nur Rechtecke proportional der Grundlinie und Höhe, sondern auch Dreiecke, und indem man jedes Polygon in Dreiecke und Rechtecke mit parallelen Grundlinien und Höhen auflösen kann, läßt sich jedes Polygon in ein anderes verwandeln, bei dem alle Grundlinien oder alle Höhen im gleichen Verhältnis verändert sind. Leicht schätzbar sind aber nur Rechtecke und allenfalls Dreiecke.

Zeichnet man Rechtecke mit gleicher Grundlinie, so ist ihre Fläche der Höhe proportional. Sie können dann nur eine Größe darstellen, ebenso wie die geraden Linien, und haben dieser gegenüber den Nachteil, daß sie die Anordnung auf der Abszissenachse weniger scharf erkennen lassen, und daß für eine etwa erwünschte Verbindungslinie ihrer Endpunkte weniger Raum übrig bleibt. Dagegen bieten sie den Vorteil, daß sie sich besser in Teile einteilen lassen, die durch Farbe oder Schraffierung deutlich zu unterscheiden sind. So kann durch eine Reihe von Rechtecken gleicher Basis die Summe oder Differenz zweier Reihen veranschaulicht werden. Die gleiche Möglichkeit besteht zwar auch für Liniendiagramme, bei diesen kommt es aber leicht zu einer optischen Täuschung, die bei Summierung vieler Reihen schwer vermieden werden kann. Gibt man nämlich auf einer Figur die Kurve einer Reihe und ihrer Summe mit einer zweiten, dann diese Summe mit einer dritten, vierten usw., so entsprechen die Abstände auf den Ordinaten den jeweils hinzukommenden Reihen. Jede folgende Kurve hat also zu ihrer Grundlage nicht die gerade Grundlinie, sondern die vorhergehende Kurve. Dem Auge drängen sich nun nicht die senkrechten Abstände auf den Ordinaten, sondern die kürzesten schrägen Abstände der Kurventeile auf, und diese sind verhältnismäßig umso schmäler, je steiler die untere Kurve an der betreffenden Stelle verläuft. Die oberen addierten Reihen werden also in verzerrten Kurven wiedergegeben und sind schwer erkennbar. Betont man dagegen die Ordinaten und verbreitert sie zu Rechtecken, so erscheinen auch die addierten Stücke in ihrem richtigen Verhältnis. Allerdings ist dadurch der Zusammenhang zerrissen. Um diese Mängel zu vermeiden, könnte man die Rechtecke durch Zwischenräume trennen, und in diesen die Verbindungslinien ziehen, oder in einer reinen Kurvenfigur alle einzelnen Summanden auf der Grundlinie beginnen lassen, dafür aber die unwichtigeren Zwischensummen fortlassen. Beide Lösungen können leicht zur Übersichtlichkeit führen. daher ist in jedem Fall zu untersuchen, welche Darstellungsform die zweckmäßigste ist.

Die Unterteilung ist übrigens nicht auf Rechtecke von gleicher Grundlinie beschränkt, sondern bei allen Rechtecken gleichmäßig durchführbar. Sie kann auch in der Form geschehen daß die Rechtecke in kleine Quadrate oder Rechtecke von gleicher Form geteilt werden, die sich leicht zählen lassen, um dadurch die Abschätzung zu erleichtern. Die Unterteilung von Dreiecken ist weniger zu empfehlen. Teilt man die Grundlinie und verbindet die Schnittpunkte mit der Spitze, so sind zwar die entstehenden kleinen Dreiecke den Abschnitten der Grundlinie proportional, aber unter sich weniger leicht zu vergleichen, weil die mittleren kürzer und breiter, die äußeren länger und schmäler erscheinen. Teilt man die Höhe ein und zieht Parallelen zur Grundlinie, so entsteht ein Dreieck und Trapeze, die dem Produkt aus der Höhe und dem Durchschnitt, d. h. der halben Summe der Grundflächen proportional sind. Solche Figuren sind sowohl schwierig zu berechnen und zu zeichnen, als zu

schätzen und nachzuprüfen. Am leichtesten ist die Schätzung, Nachprüfung und Vergleichung der Teile außer bei Rechtecken bei Quadraten und Kreisen und zwar bei letzteren, wenn sie in Sektoren zerlegt werden; schwerer ist die Schätzung des Inhalts von Ringen.

Quadrate, Kreise und überhaupt alle unter sich ähnlichen Figuren haben die Eigenschaft, daß ihre Flächeninhalte sich verhalten wie die zweiten Potenzen entsprechender Längen. Sie eignen sich daher zur Darstellung von Reihen, von denen die eine aus Quadraten oder Quadratwurzeln der anderen besteht. Aus diesem Grunde nimmt die Länge und Breite der Quadrate, Kreise usw. weniger zu als die durch die Fläche dargestellte Größe, sie lassen sich daher auf beschränktem Raum leichter anbringen, z. B. auf Karten. Sie werden deswegen häufig angewandt; sie haben aber das Bedenken gegen sich, daß Flächeninhalte weniger leicht geschätzt werden können als Längen; und daß häufig nicht klar ersichtlich ist, ob die Längen oder die Flächeninhalte verglichen werden sollen.

Was vom Inhalt der Flächen gesagt wurde, gilt im gesteigerten Maße vom Inhalt von Körpern. Bei diesen ist es ebenfalls möglich, Höhe, Länge und Breite den Zahlen verschiedener Reihen proportional zu machen. Dann ist der Inhalt dem Produkt dieser drei Zahlen proportional. Ist die Ausdehnung in einer der drei Dimensionen konstant, so ist der Rauminhalt dem Produkt der durch die beiden anderen Dimensionen dargestellten Zahlen proportional. Sind die Körper ihrer Form nach ähnlich, so entspricht ihr Inhalt der dritten Potenz der Längenmaße. Die Unterschiede der Größe des Inhalts treten also noch weniger deutlich hervor.

Außer durch ihren Inhalt können körperliche Darstellung auch durch die Gestaltung ihrer Oberfläche wirken. Wenn nämlich ein Koordinatensystem in drei Dimensionen ausgeführt wird, nimmt es körperliche Gestalt an. Dies hat den Vorteil, eine veränderliche Größe darzustellen, die von zwei unter sich unabhängigen, veränderlichen Größen abhängig ist. Freilich ist die wirkliche Ausführung derartiger Figuren schwierig und ihre bloße Abbildung nicht leicht verständlich.[1]) Einen ungefähren Begriff kann man sich von dieser Darstellungsweise machen, wenn man auf einem Schachbrett Steine eines Baukastens aufstellt. Die Werte der einen unabhängigen Variablen zählt man von links nach rechts, die der anderen von vorn nach hinten. So entspricht jedes Feld dem Zusammentreffen von zwei unabhängigen Werten. Der in diesem Fall eintretende abhängige Wert, die Funktion wird durch die Höhe der darauf gestellten Bausteine ausgedrückt, wobei jeder der darin enthaltenen Würfel eine Einheit bedeutet. Will man Kurven in drei Dimensionen ausführen, so kann man in die Ecken der Felder Stifte von entsprechender Höhe aufstellen, die durch Fäden verbunden werden.[2])

Zur Darstellung in Buchform eignen sich solche Figuren allerdings wenig. Ein leicht anwendbarer Ersatz kann aber nach dem Vorschlag von Turquan und Körösi[3]) die Darstellung der dritten Größe statt durch die Ausdehnung in der dritten Dimension durch Nüancen der Farbe oder Schraffierung geben, ähnlich wie auf Landkarten die Niveauunterschiede in leicht verständlicher Weise durch verschiedene Farben bezeichnet werden. Zur richtigen Würdigung

[1]) Gipsmodell des Altersaufbaues einer Bevölkerung bei den aufeinanderfolgenden Volkszählungen; (für Schweden bei Gabaglio, Teoria generale della Statistica). II. hinter S. 434.

[2]) Durch Aneinanderfügen der im Anhang III erwähnten ausgeschnittenen Diagramme oder Aufstellung derselben mit den Nullpunkten auf der dritten Achse erhält man ein dreidimensionales Diagramm.

[3]) Bulletin de l'Institut International de Statistique, 1901. Tome XIII. P. 313 ff.

der Farben ist dann diese Skala an der Seite zu wiederholen, wobei die durch sie ausgedrückten Größen in dem Maßstab der Hauptfigur zu halten sind.

Zur leichteren Ausführung werden auch an Stelle der Körper häufig deren Abbildungen gegeben. An sich ist dies vollkommen zulässig, wenn eine klare unzweideutige Vorstellung erweckt wird, daß Körper verglichen werden sollen, und deren Inhalt den dargestellten Zahlen proportional ist. Da sich die Inhalte ähnlicher Körper jeder Form wie die dritte Potenz von entsprechenden Längen verhalten, können diese Bilder auch Gegenstände darstellen, z. B. Kisten, Ballen, sogar Tiere und Menschen. Nur muß dann deutlich gesagt werden, daß der Inhalt dieser Körper den dargestellten Größen entspricht. Andernfalls kann der Beschauer auf den Gedanken kommen, statt der Körperinhalte die Oberfläche der Bilder auf dem Papier oder entsprechende Längen zu vergleichen und als proportional der dargestellten Zahl anzusehen. Diese Gefahr wird dann verstärkt, wenn außer ähnlichen Körpern auch Reihen gleicher Körper, wie Geldrollen, dargestellt werden, bei denen die Längenmaße das entscheidende sind. Am schlimmsten wird der Fehler, wenn die Zeichnung selbst nicht den Inhalt des Körpers, sondern dessen Länge oder Oberfläche der Zahl proportional macht, womöglich dies nicht deutlich hervorgeht oder gar zwischen Körperflächen und Längenmaßen hin- und herschwankt.

(Vergl. Prof. Hickmanns Geographisch-Statistischer Universal-Taschenaltlas; die Tafel über Staatsschulden hat schwankenden Maßstab, die Tafel über Ein- und Ausfuhr Flächenmaße, die Tafel über Schiffsräume dagegen Hohlmaße.)

Will man körperliche Figuren korrekt bildlich wiedergeben, so müßte man genau genommen stereoskopische Aufnahmen von ihnen machen; d. h. man müßte die Projektionen ihrer Linien auf eine zwischen sie und die Augen gehaltene Ebene für jedes der Augen besonders zeichnen. Die Projektionen weichen ein wenig voneinander ab, da das rechte Auge den Gegenstand ein wenig mehr von rechts, das linke mehr von links erblickt. Die gleichzeitige Betrachtung der durch den Apparat getrennten Bilder bringt dann den Eindruck der Körperlichkeit hervor.

Die übliche perspektivische Zeichnung ist bereits ein Kompromiß. Anstelle der Bilder zweier Augen wird nur das Bild eines Auges gezeichnet. Immerhin erfordert eine richtige Perspektive, zeichnerisches Können oder eine geometrische Konstruktion, deren Erfolg in keinem Verhältnis zur aufgewendeten Mühe steht. Im Gegenteil, da die Seiten und Oberflächen nur sichtbar werden, wenn die Figur nicht genau von vorne betrachtet wird, werden alle Linien verkürzt, und zwar in verschiedenem Maße, die entfernteren stärker als die näheren. Die Nachmessung ist daher sehr erschwert.

Noch stärker tritt dieser Übelstand bekanntlich bei der Photographie auf, die sonst ein willkommenes Hilfsmittel wäre. Perspektivische Zeichnungen und Photographien sind daher nur bei einfachen Figuren, wie Würfel und Zusammenstellungen von solchen ausführbar, die leicht zu zeichnen sind, und bei denen nur eine Kante nachgemessen oder verglichen zu werden braucht, um den Inhalt zu berechnen.

Eine genauere Nachmessung gestattet die Parallelprojektion, wie sie annähernd durch den Sonnenschatten bewirkt sind. Da aber die Sonne kein mathematischer Punkt ist, sind ihre Schatten nicht scharf, und daher ebenfalls zur Nachbildung unbrauchbar. Auch hat eine Parallelprojektion den Nachteil, daß, um die Seiten- und Oberansicht gewinnen zu können, der Gegenstand nicht genau von vorne aufgenommen werden kann. Dadurch wird auch die Vorderansicht etwas verzerrt.

Am bequemsten und für praktische Zwecke genügend ist eine allerdings ungenaue Darstellungsweise, welche die Vorderseite des Körpers im Aufriß, also genau von vorne, gibt und daran die rechte (bezw. linke) und obere Ansicht in Parallelprojektion anschließt. Sie ist leicht auszuführen und nachzuprüfen und doch noch verständlich.

Endlich kann man auch den Aufriß der Vorderansicht mit perspektivischer Darstellung der Seiten- und Oberansicht verbinden. Eine Nachmessung ist dann allerdings nur für die Vorderansicht möglich und diese Form eignet sich deshalb nur für Würfel und andere gleichgestaltete Figuren mit gleichem Abstand vom Beschauer oder für Körper mit gleicher Tiefe, deren Inhalt ihrer vorderen Fläche proportional ist, und die nur zur Belebung des Bildes körperlich gezeichnet werden sollen.

Faßt man die bisher dargestellten geometrischen Formen nach dem Zweck zusammen, so ergibt sich folgende Zusammenstellung:

Bei der unmittelbaren Vergleichung von Größen, also bei der Veranschaulichung einer Zahlenreihe bietet sich als Mittel das Liniendiagramm, die Reihe von Rechtecken mit gleicher Grundlinie und verschiedener Höhe (oder auch mit gleicher Höhe und verschiedener Grundlinie), Reihen von Quadraten, Kreisen und anderen regelmäßigen Figuren mit proportionalen Flächen oder Körper und Bilder von solchen mit proportionalem Rauminhalt. Das Liniendiagramm ist, wie bereits ausgeführt, besonders geeignet, die Reihe als Funktion einer konstant zunehmenden Größe, z. B. der Zeit darzustellen; Flächen und Körper bieten Abwechslung und ein lebendigeres Bild und haben die Eigentümlichkeit, die Unterschiede der Längenmaße auf ein geringeres Maß, die Quadrat- bezw. Kubikwurzel herabzusetzen. Rechtecke und Dreiecke mit veränderlicher Grundlinie und Höhe stellen zwei Reihen und die Produkte ihrer Glieder dar, sind aber weniger geeignet durch ihre Anordnung die Beziehung zur Zeit oder einer konstant sich verändernden Größe auszudrücken. Insoweit hängt die Zweckmäßigkeit der Darstellung von der geometrischen Konstruktion ab.

4. Mannigfaltigkeit, nicht nur durch Verschiedenheit der Konstruktion, sondern auch der Maßstäbe

So mannigfaltig die geometrischen Konstruktionen auch sind, in denen die Diagramme ausgeführt werden können, so wird doch deren Vielgestaltigkeit dadurch noch weit gesteigert, daß sie bei Wahl verschiedener Maßstäbe verschiedene Größe zeigen, und, wenn in derselben Figur zwei oder mehr unabhängige Maßstäbe zu wählen sind, auch ihre Form sehr verschieden ausfallen kann.

Diese reiche Auswahl von Ausdrucksmitteln erleichtert es zwar einerseits, für jeden Zweck die geeignete Form zu finden, sie erschwert aber auch die Vergleichung. Es ist daher im einzelnen Fall zu untersuchen, ob bloße Anschaulichkeit oder auch Vergleichbarkeit erstrebt wird, und auf was sich die Vergleichung beziehen soll. Genau genommen stellt jede graphische Darstellung einer Reihe von Größen schon eine Vergleichung derselben unter sich dar, im eigentlichen Sinne wird die Vergleichbarkeit aber erstrebt, wenn mehrere Diagramme mit einander in Beziehung gesetzt werden sollen.

Die Wahl des passenden Maßstabes ist eine rein praktische Frage, wenn nur ein Maßstab in einer Figur oder einer Reihe von Figuren anzuwenden ist. Werden Größen durch Linien, Quadrate, Kreise, Würfel oder durch Rechtecke von gleicher Grundlinie dargestellt und drückt der Abstand der Figuren von

einander und die Grundlinie der Rechtecke keine Zahlen aus, so ist die Wahl des Maßstabes ohne Einfluß auf die Form. Es ist nur darauf zu achten, daß die Figuren wirklich in demselben Verhältnis zur Einheit stehen wie die Zahlen; ferner daß sie möglichst weder über den verfügbaren Raum herausreichen und daher abgebrochen werden müssen, noch daß sie zu klein und undeutlich werden.

Anders ist es bei Figuren, in denen zwei oder mehr von einander unabhängige Maßstäbe vorkommen: bei Rechtecken mit verschiedenen Grundlinien und Höhe, bei Punkten, Linien, Flächen und Körpern, deren Lage im Raum eine Zahlenbeziehung ausdrückt, also vor allem bei den in ein Koordinatensystem eingezeichneten Kurven.

Bei diesen wird nicht nur die Größe von der absoluten Länge der Maßstäbe, sondern auch die Gestalt der Figuren von dem Verhältnis der Maßstäbe zu einander bestimmt. Ein Rechteck, dessen eine Seite z. B. der Bevölkerungszahl, dessen andere der Schuldenlast pro Kopf proportional ist, kann sich einem Quadrat oder einem schmalen Streifen nähern, ohne daß die eine oder andere Form an sich unrichtig wäre. Ebenso können im Koordinatensystem steile und flache Kurven eingezeichnet werden, die denselben Gegenstand gleich richtig in dem betreffenden Maßstab ausdrücken. Werden in einem Diagramm zwei oder mehrere Kurven verschiedenartiger Gegenstände eingezeichnet, so ist ein gemeinsamer Maßstab nicht gegeben und der Vergleich der Kurven hängt davon ab, wie die Maßstäbe gewählt worden sind. Dieser Übelstand hat zu mancherlei Versuchen und Vorschlägen zur Abhilfe geführt und das Internationale Statistische Institut hat sich auf seinen Sitzungen im Jahre 1901 und 1909 eingehend mit dieser Frage der Vergleichbarkeit beschäftigt, ohne zu einem abschließenden Ergebnis zu gelangen.

Während Bertillon[1]) das Hauptgewicht auf die Anschaulichkeit und leichte Verständlichkeit legte, und die Wiedergabe der beobachteten Größen selbst empfahl, betonte Lucien March[2]) die Wichtigkeit der Vergleichbarkeit und befürwortete eine Umgestaltung der Kurven in der Weise, daß die Verschiedenheit der absoluten Zahlen nicht den Vergleich hindern könne. Er schlug dazu Kurven der Logarithmen, der relativen Zu- oder Abnahme, und vor allem der Verhältniszahlen der einzelnen Glieder zu einem aus jeder Reihe abgeleiteten Mittelwert vor, der durch eine für alle Reihen gleiche Länge der Ordinaten auszudrücken sei.

Diese Verschiedenheit der Vorschläge beruht anscheinend darauf, daß verschiedene Zwecke mit den Diagrammen angestrebt werden. Sollen sie vor allem zur Veranschaulichung des Zahlenmaterials dienen, so ist die Wiedergabe der absoluten Zahlen unentbehrlich. Gleichartige Größen sind soweit als möglich mit gleichartigem Maßstab darzustellen, bei ungleichartigen ist das Verhältnis der Maßstäbe beliebig und so zu wählen, daß die größtmöglichste Deutlichkeit erreicht wird. Man wird also danach streben, den verfügbaren Raum so gut als möglich auszunutzen und weder durch zu große Steilheit noch zu große Flachheit der Kurve die Anschaulichkeit zu beeinträchtigen. Dieser Zweck der Anschaulichkeit führt bei populären Darstellungen zu einer großen Freiheit in der Wahl des Maßstabes; bei wissenschaftlichen Werken ist darauf zu achten, daß die Maßstäbe in demselben Werk tunlichst beibehalten werden, um eine Vergleichung zu ermöglichen. Immer wird aber das Hauptgewicht darauf gelegt, daß die beobachteten Größen selbst und nicht erst Ableitungen von ihnen dargestellt und verglichen werden.

[1]) Tome XIII, Vol. 1, P. 132, 313. Propositions rélatives à l' uniformirté à apporter dans l'établissement des graphiques.

[2]) Tome XIX. S. 118*, 50 Les moyens de rendre comparahles les courbes statistique.

Durch die Wiedergabe der ursprünglichen Zahlen wird nun aber eine genauere Vergleichung der Kurven oft erschwert. Wenn die Größen der einen Reihe beständig viel größer als die einer anderen sind, so bewirken relativ geringe Veränderungen in der ersten Reihe eine große Steilheit in der ersten Kurve, während relativ gleiche Veränderungen in der zweiten Reihe die Kurve kaum wahrnehmbar beeinflussen. Es tritt so das Bedürfnis auf, die Kurvengestaltung vor den absoluten Größen unabhängig zu machen. Dies kann dadurch erreicht werden, daß für jede Kurve der Durchschnittswert aller Glieder, oder der Wert des Anfangs-, Endgliedes oder des Gliedes für einen gemeinsamen Zeitpunkt gleich derselben Länge der Ordinaten gesetzt wird, und diese zu dem Maßstab der Abszissen in einem festen Verhältnis steht.

Ferner können die Schwankungen an sich, ohne Beziehung auf die Größe der Glieder Gegenstand des Interesses sein, oder die Beziehungen der Glieder zu einander, sei es die Differenzen und Quotienten aufeinanderfolgender Glieder oder von Gruppen. So entstehen vielerlei Gesichtspunkte, unter denen Kurven verglichen und zu diesem Zwecke umgestaltet werden können.

Die Vorschläge zur größeren Vergleichbarkeit der Kurven in diesem Sinne gehen von dem Grundsatz aus, daß es zulässig ist, die Einheit der Länge nicht der Zahl 1, sondern einer anderen Zahl gleichzusetzen, oder was dasselbe bedeutet, die ursprüngliche Zahl zu dividieren, um darstellbare Zahlen zu erhalten. Sie erweitern diesen Grundsatz, indem sie andere Rechenoperationen an den Gliedern der Reihen vornehmen und die veränderte Reihe in Kurvenform zeichnen.

Sie bezwecken damit

1. Beseitigung der Willkür in der Wahl der Maßstäbe,
2. größere Übereinstimmung von Kurven, wenn die feinen Unterschiede beobachtet werden sollen,
3. selbständige Darstellung der Beziehungen der Größen einer Reihe untereinander, und zu denen anderer Reihen.

Die vorgeschlagenen und angewandten Rechenoperationen sind: Division, Subtraktion, Addition, Durchschnittsberechnung, Potenzieren, Logarithmieren usw. einzeln und in Verbindung und Wiederholung, und zwar zwischen Gliedern derselben und mehrerer Reihen.

Der einfachste Zweck ist die Veranschaulichung einer Reihe von Größen, z. B. einer zeitlichen Entwicklung, wie die der Temperatur eines Fieberkranken. Hier ist jeder Maßstab für Ordinaten und Abszissen richtig, wenn er konsequent durchgeführt wird. Es ist nur darauf Rücksicht zu nehmen, daß die gewählten Einheiten sich durch bekannte Längen, z. B. 1 cm, leicht nachmessen lassen, und daß die entstehende Figur sich dem Raum anpaßt, und wenn erwünscht, eine Fortsetzung ermöglicht. Benützt man kariertes Papier, so ergibt sich der passende Maßstab aus dessen Größe und Einteilung von selbst.

Außer zur Veranschaulichung dienen die Diagramme aber auch zur leichteren Vergleichung von Zahlenreihen. Hierfür sind besonders die Koordinatensysteme den Ziffern weit überlegen; während man nur wenige Ziffern gleichzeitig wahrnehmen und gegeneinander abwägen kann, übersieht das Auge leicht zwei Kurven und schätzt nicht nur die Höhe der Ordinaten, sondern auch die Steilheit oder Flachheit der Verbindungslinien und die Ähnlichkeit oder Unähnlichkeit der Verlaufs. Auf diese Vergleichbarkeit ist deshalb besonderer Wert zu legen, und zwar auch dann, wenn auf einer Figur nur eine Kurve dargestellt ist, diese aber mit Kurven gleichartiger Gegenstände, die außerdem gegeben werden, oder bekannt sind, verglichen werden soll.

Soll sich der Vergleich sowohl auf die absolute Größe der Zahlen, als auf deren Zu- oder Abnahme beziehen, so muß für gleichartige Größen ein einheitlicher Maßstab gewählt werden, und zwar sowohl für Ordinaten als für Abszissen. Das Verhältnis beider Maßstäbe zu einander ist so zu wählen, daß die ganze Figur übersichtlich und leicht herzustellen und nachzuprüfen ist.

Sind die absoluten Größen zweier Reihen sehr verschieden, z. B. die eine tausendfach größer als eine Zahl der anderen Reihe, so sind sie zwar trotzdem, soweit möglich auf demselben Papier zu vereinigen, außerdem aber die kleineren Größen in einer besonderen Figur mit größerem Maßstab anschaulich und erkennbar zu machen. Dieser größere Maßstab sollte zu dem der Hauptfigur in einem einfachen Verhältnis stehen, z. B. 1 : 2, 1 : 10, 1 : 100 usw.

Wird das Verhältnis der absoluten Größen als bekannt oder nicht interessierend nicht dargestellt, oder handelt es sich um Reihen verschiedenartiger Größen, so ist kein gemeinsamer Maßstab der Ordinaten vorhanden. Will man trotzdem eine Willkür ausschließen, so muß man die Maßstäbe nach einem einheitlichen Prinzip wählen.

Ist in zwei oder mehreren Koordinatensystemen je eine der darzustellenden Größen gleichartig, bedeutet z. B. die Länge der Abszissen die Zeit, so kann man eine Vergleichbarkeit der Ordinaten dadurch erzielen, daß man gemeinsame Berührungspunkte herstellt. Man wählt zu diesem Zweck die Maßstäbe der Ordinate so, daß die Ordinaten für eine oder für zwei bestimmte Abszisse gleich werden. Diese Ordinaten können am Anfang, am Ende oder einer beliebigen Stelle stehen, sie müssen nur in allen Figuren vorkommen. Um die Willkür noch mehr anzuschließen, kann man den Durchschnittswert einer Anzahl entsprechender Glieder jeder Reihe durch gleiche Ordinate ausdrücken.

Am einfachsten für die Rechnung ist es, die entscheidenden Werte gleich 100 zu setzen. Um zu einer gleichmäßigen Grundlage zu gelangen, empfiehlt L. March (Bulletin de l'Institut de St. Tome XIX, S. 56) in Anlehnung an die von englischen Board of Trade berechneten Indexziffern den Durchschnitt der Jahre 1891—1900 gleich 100 zu setzen, soweit Daten für diese Zeit vorliegen.

Als Maßstab der Abzissen schlägt L. March vor, den Abstand für 30 Jahre gleich der Ordinate für 100 zu setzen. Damit wäre ein Format erreicht, das sowohl für einen Zeitraum von 100 Jahren als Maximum der Beobachtungszeit wie für 10 Jahre als Minimum fruchtbarer Vergleichung noch handlich wäre. Das Verhältnis 30 : 100 hat nun den Nachteil weder im Dezimalsystem runde Zahlen zu ergeben, noch der üblichen Einteilung des Papieres in cm und mm angepaßt zu sein. Es ist auch leichter eine Strecke zu halbieren und dann in Viertel zu teilen, als in Drittel. Berücksichtigt man, daß Reihen über einen Zeitraum von 100 Jahren sehr selten sind, und bei größeren Kurven doch ein größeres Format gewählt werden muß, was sich bequemer in ein Buch einfügen läßt, wenn es nur nach der Seite entfaltet zu werden braucht, so scheint es zweckmäßiger, die Abszisse für 20 Jahre, höchstens 25 Jahre der Ordinate für 100 zu setzen. (Vergl. Roesle a. a. O.) Noch praktischer wäre vielleicht 10 Jahre zu wählen, wobei auch eine Steigerung der Ordinaten um 20, 30 und mehr nicht zu steil wird und ein Maximum von 200 oder 300 in dem normalen Rahmen bleibt.

So wertvoll die Umgestaltung statistischer Reihen in Prozentzahlen und deren Darstellung in Diagrammen auch ist, ihre allgemeine Anwendung wäre nicht wünschenswert. Denn die zugrundeliegenden absoluten Zahlen werden vollständig unerkennbar. Damit wäre ein wichtiger Vorteil der graphischen Darstellung und besonders des Kurvendiagrammes verloren, nämlich die Möglichkeit, die Glieder einer Reihe unmittelbar mit denen einer anderen vergleichen zu können. Indexziffern und deren Kurven haben ihre selbständige Bedeutung, können aber nicht die ursprünglichen Reihen und Kurven ersetzen.

Da gleiche gilt von anderen abgeleiteten Werten, die in der Weise ausgedrückt werden, daß der Maßstab in entsprechender Weise gewählt wird. Man kann durch geeignete Bemessung der Maßstäbe erreichen, daß die Kurve im allgemeinen in einem Winkel von 45° verläuft, daß die durchschnittliche Differenz oder der durchschnittliche Quotient aufeinanderfolgender Ordinaten durch eine gleiche Länge wiedergegeben werden, die auch dem Abstand der Abszissen gleich ist.

Hierher gehört der Vorschlag eines ungenannt gebliebenen Mitglieds der Kommission des Internationalen Statistischen Instituts (Band XIII, Seite 315), den Maßstab so zu wählen, daß eine „allure moyenne" der Kurve von 45° entsteht. Es ist nicht genau ersichtlich, ob damit gemeint sein soll, daß jedes Stück der Kurve eine solche Steilheit besitzen soll, daß deren Durchschnitt 45° beträgt, oder, wie es leichter konstruktiv darzustellen wäre, daß der Gesamtverlauf der Kurve als Einheit gedacht, in einem Winkel von 45° verlaufen soll. Dann würde die Abszissenachse gleich der Differenz des ersten und letzten Gliedes zu machen sein; im ersteren Falle wären die Winkel zu messen und aus ihnen der Durchschnitt zu bilden, oder etwa die durchschnittliche Differenz der Ordinaten gleich ihrem Abstand zu machen. In beiden Fällen wäre gerade die charakteristische Eigentümlichkeit der Kurve, die Steilheit oder Flachheit im Ganzen bezw. in den einzelnen Teilen auf einen nichtssagenden überall gleichen Durchschnittswert zurückgeführt, und nur mehr zu erkennen, ob die Steilheit in dem einen Teil der Kurve größer, im anderen kleiner ist, als in dem Mittelwert von 45°. Zudem würden von Natur flache Kurven ein außerordentlich hohes Format erhalten, falls nicht der untere Teil weggelassen wird.

5. Reihen von Beziehungszahlen zwischen den Gliedern einer oder mehrerer Reihen

Bisher handelte es sich darum, die Reihen von Zahlen selbst zur Darstellung zu bringen, und durch Wahl geeigneter Maßstäbe vergleichbar zu machen. Bei der Betrachtung der Figuren entsteht dann zugleich das Bild der Zu- oder Abnahme innerhalb jeder Reihe. Will man diese genauer untersuchen, so kann man die Zu- oder Abnahme selbst zum Gegenstand der Darstellung machen, und zwar entweder die absolute Zunahme, d. h. die Differenzen der aufeinanderfolgenden Glieder, oder die relative Zunahme, d. h. die Quotienten jedes Gliedes durch das vorhergehende Glied.

Es ist üblich, diese Differenzen und Quotienten, letztere in Prozent ausgedrückt, in Form von Kurven oder Verbindungslinien der Ordinaten zu zeichnen.

Dagegen ist eingewendet worden,[1]) daß es in diesem Fall der Logik wider-

[1]) Bertillon (Bulletin de l'Inst. int. de Stat., Tome XIII, P. 316, § 9). beurteilt die abgeleiteten Kurven, unter Hinweis auf das Beispiel einer Kurve der Differnzen, abfällig und sagt:

„§ 9: Si, au lieu de cela, on portait sur des ordonnées equidistantes dont l'equidistance serait également celle d'une année, des longueurs proportionelles aux quantités dont la taille s' est accrue pendant l'année considérée, on aurait encore une représentation graphique du phénomènc; mais elle n'aurait plus le même interêt. En effet la courbe passant par ces points n' aurait plus la même signification que ci dessus; un point intermédiaire de la courbe ne représenterrait rien de simple, et la continuité graphique établie par le tracé n' ajouterait rien aux indications données par les points isolés, qui seuls exprimcraient le phenomène. On n'aurait donc rien gagné à relier ces points par une courbe."

streite, die Endpunkte der Ordinaten zu verhinden. Werde nämlich durch die Ordinaten die wachsende Größe selbst dargestellt, so könne man die Abszisse, also die Zeit, in beliebig viele Teile teilen, und erhalte durch die an jedem Punkt errichtete Ordinate ein richtiges Bild der vermutlichen Größe. Die Verbindungslinie der Endpunkte der Ordinaten stellt so eine Interpolation dar, die mit der Wirklichkeit nicht im Widerspruch steht, sondern vielmehr ihr näher kommt, wenn anzunehmen ist, daß die Entwicklung kontinuierlich verläuft.

Für die Differenzen der aufeinanderfolgenden Glieder sei dagegen eine solche Interpolation unzulässig. In der Tat bietet die Interpolation bei Bewegungsmassen Schwierigkeiten. Zunächst könnte man daran denken, daß die Zunahme umso kleiner wird, je kleiner die Zeiträume werden. Hat man die Ordinaten für die Zunahme innerhalb ganzer Jahre gezeichnet, so würde für halbe Jahre durchschnittlich der halbe, für Vierteljahre der vierte Teil des Wertes sich ergeben. Eine kontinuierliche Reihe, die unendlich kleine Zeiträume darstellt, würde die Zunahme unendlich klein werden lassen. Diesem Bedenken ist zu entgegen, daß man die Zunahme stets auf ein ganzes Jahr berechnen kann.

Erheblicher ist jedoch folgender Einwand:

Berechnet man die Zunahme für ganze Jahre, so stellen die gefundenen Zahlen Durchschnittswerte der Zunahme der einzelnen Teile jedes Jahres dar. Nimmt man an, daß die Zunahme innerhalb jedes Jahres unverändert bleibt, und dann plötzlich in diejenige des nächsten Jahres übergeht, so erhält man eine treppenförmige Reihe von Rechtecken, deren Höhe der Zunahme der betreffenden Jahre proportional ist. Verbindet man die Mittelpunkte der oberen Seiten dieser Rechtecke, so hat man eine Linie, die für jeden Zeitpunkt den Durchschnittswert für die Jahreszeitstrecke angibt, deren Mitte er bildet, unter der Voraussetzung, daß die Zunahme innerhalb der Beobachtungsjahre konstant ist. Es wird jedoch der Anschein erweckt, als ob die Zunahme für jeden kleinsten Zeitraum dargestellt werden sollte.

Diese Vorstellung steht mit der Wirklichkeit im Widerspruch. Sie würde erfordern, daß die von der Kurve und den sie begrenzenden Ordinaten eingeschlossene Fläche gleich dem Rechteck zwischen denselben Ordinaten ist. Dies ist jedoch nur ausnahmweise der Fall, wenn nämlich das vorhergehende Rechteck um denselben Betrag kleiner ist, als das nachfolgende größer oder umgekehrt. Am größten ist die Abweichung bei einem Maximum oder Minimum der Zunahme bezw. Abnahme.

Wächst z. B. die Bestandsmasse in einem Jahre um 10 Einheiten, im nächsten um 12, dann wieder um 8, so wird das mittlere Rechteck 12 Flächeneinheiten enthalten. Zieht man die Verbindungslinien der Mittelpunkte, so werden von diesem Rechteck 2 Dreiecke abgetrennt, von denen das eine $^1/_2$, das andere $^2/_2$ jener Flächeneinheiten enthält, die Gesamtfläche wird also von 12 auf $10^1/_2$ Einheiten vermindert, das ist um $^1/_8$ ihrer Größe.

Eine korrekte Interpolation müßte die Kurve noch über 12 ansteigen lassen, denn nur dann kann sich als Durchschnitt für ein Maximumjahr der Wert 12 ergeben.

Eine solche Interpolation ist nun aber nur durch komplizierte Rechnung oder Konstruktion durchführbar.[1]) Daher empfiehlt es sich, für praktische Zwecke, wenn Genauigkeit angestrebt werden soll, Zu- und Abnahme, d. h. Bewegungsmassen, durch Rechtecke auszudrücken.

Eine solche Genauigkeit ist dann wünschenswert, wenn Bewegungsmasse und Bestandmasse gleichzeitig dargestellt und ihr Zusammenhang genau wieder-

[1]) Siehe Anhang: Graphische Ausgleichung bezw. Interpolation einer Reihe durch Parabeln.

gegeben werden soll. Wird die Bestandmasse durch gradlinige Verbindungslinien der Ordinaten wiedergegeben, so entspricht jeder Punkt dieser Verbindungslinie dem Resultat der Fläche der Bewegungsmasse, die durch Rechtecke ausgedrückt ist.

Werden dagegen die Mittelpunkte der oberen Seiten dieser Rechtecke verbunden, so tritt diese Übereinstimmung nur für die diesen Punkten entsprechenden Ordinaten ein. Genau genommen müßte jedem gradlinigen schrägen Stück der Linie der Bewegungsmasse ein Parabelstück für die Bestandmasse entsprechen.

In der Praxis werden allerdings oft sowohl die Urzahlen als ihre Differenzen durch gradlinig verbundene Ordinaten dargestellt, so daß die erwähnte Ungenauigkeit eintritt. Dieselbe kann dadurch auf ein möglichst geringes Maß zurückgeführt werden, daß man die Ordinaten der Differenzen in die Mitte zwischen die dazugehörigen Ordinaten der Urzahlen stellt, und am Anfang und Ende horizontale oder sonst unter sich parallele Strecken von der Hälfte des Abstandes der Ordinaten hinzugefügt. Dann gleichen sich die Fehler für die ganze Reihe wieder aus. Läßt man z. B. für das erste halbe Jahr die Linie der Bewegungsmasse horizontal verlaufen und verbindet erst dann die Mittelpunkte der oberen Seiten, so stimmt für diese Mittelpunkte die Ordinate der Bestandsmasse mit dem Resultat der Bewegungsmasse überein; denn die zugefügten und hinweggenommenen Dreiecksflächen gleichen sich an diesen Punkten genau aus.

Man erhält so genaue Wiedergabe des Durchschnitts zwischen zwei Erhebungsdaten der Bestandmasse und zu den Beobachtungsterminen für die Bewegungsmasse. Beide Kurven sind also ein wenig ausgeglichen, der Einfluß der Oscillationen ist vermindert, was ihren Wert für praktische Zwecke eher vermehren dürfte.[1])

Statt der Differenzen kann man auch die Quotienten der aufeinanderfolgenden Glieder ins Auge fassen. Damit nähert man sich besonders dann den Bedingungen der Wirklichkeit, wenn die Zunahme einer Größe von dem jeweils erreichten Bestand abhängt, also bei konstanter Zunahme eine geometrische Progression entstehen würde. Ein bekanntes Beispiel ist das Anwachsen eines auf Zinseszins angelegten Kapitals bei gleichbleibendem Zinsfuß, oder das einer Bevölkerung bei gleichbleibenden Prozenten des Geburten-Überschusses.

Korrekter Weise müßte man in solchen Fällen die Zinseszinsformel $K_n = K_o \left(\frac{100 + p}{10}\right)^n$ anwenden, wobei K_o das Anfangskapital, K_n das Endkapital, p den Zinsfuß und n die Zahl der Jahre bedeutet. Solange aber n und p nicht zu groß werden, reicht $K_o \left(\frac{100 + p}{100}\right)^n$ nicht stark von dem leichter zu berechnenden $K_o \left(\frac{100 + np}{100}\right)$ ab. Für n = 1 lauten beide Formeln $K_n = K_1 = K_o \left(\frac{100 + p}{100}\right)$. Setzt man $1 + \frac{p}{100} = q$, so ist q der Quotient des Gliedes durch das vorhergehende. Umgekehrt ist $p = 100\,(q - 1)$.

Die Kurve der Quotienten in 100fachem Maßstab ausgeführt, unterscheidet sich also von der Kurve der Zuwachsprozente nur dadurch, daß die Null-Linie verschoben ist. Sie gibt eine richtigere Vorstellung von der Bedeutung der Zunahme, beansprucht aber mehr Raum.

(Nicht zu verwechseln mit dieser Reihe der Quotienten aufeinanderfolgender Glieder ist natürlich die Reihe der Quotienten jedes Gliedes durch das Anfangsglied; diese ist einfach eine Reduktion der Reihe, ähnlich den Indexziffern und beträgt $^1/_{100}$ der letzteren.)

[1]) Siehe Anhang IV.

Ebenso, wie man die ursprünglichen Kurven leichter vergleichbar macht, indem man die Durchschnittswerte oder Werte entsprechender Glieder = 100 setzt, kann man auch Reihen von Differenzen leichter vergleichen, indem man ihren Durchschnittswert = 100 setzt, oder sonst durch dieselbe Ordinate ausdrückt.

Man ersieht dadurch, ob sich die Zunahme über die Zeit in ähnlicher oder unähnlicher Weise verteilt. Das Resultat ist dasselbe, als wenn man den Durchschnitt der Zahlen = 100 gesetzt und die Differenzen dargestellt hätte.

Die durchschnittliche Differenz ist gleich der Differenz des ersten und letzten Gliedes, dividiert durch die Zahl der Differenzen, weil man durch Summierung der Differenzen die Reihe wieder erhält, wenn man sie zu dem ersten Glied hinzufügt.

Wollte man dasselbe Verfahren bei den Quotienten ausführen, so würde man das erste Glied mit den Quotienten der Reihe nach zu multiplizieren haben. Dies läßt sich nur rechnerisch, nicht graphisch durch Konstruktion ausführen, außer durch eine umständliche Verwandlung von Rechtecken. Sind die Verhältnisse der Glieder zu einander nicht durch Quotienten, sondern durch Prozente der Zu- und Abnahme ausgedrückt, so könnte man versucht sein, die Prozentsätze zu addieren. Man erklärt aber dann ein falsches Bild.

Eine Zunahme von 20% wird durch eine Abnahme von 20% nicht nur aufgehoben, sondern überwogen, denn $\frac{100 + 20}{100} \cdot \frac{100 - 20}{100} = 1{,}2 \quad 0{,}8 = 0{,}96$. Um die Zahl 1 wieder zu erhalten, muß man die Abnahme nicht von 100, sondern auf 100 rechnen; also multiplizieren $\frac{100 + 20}{100} \cdot \frac{100}{100 + 20} = 1$.

6. Logarithmische Kurven

Ein weiterer Fehler kann aber nicht so leicht beseitigt werden. Eine zweimalige Zunahme von z. B. 20% ist nicht gleich einer einmaligen von 40%, sondern von 1,2 . 1,2 = 1,44 = 44%.

Um die relative Zu- und Abnahme in addierbarer Form auszudrücken, kann von dem Beispiel des Zinseszinses und Diskontos ausgehen. Beim Zinsfuß von 4% = $\frac{n}{100}$ wird das Kapital jährlich um $\frac{100 + p}{100} = q$ vermehrt, in n Jahren um q^n; ich kann also angeben, in wieviel Jahren 100 zu 120, 200 usw. werden, umgekehrt vor wieviel Jahren 100 nur 80, 20, 50, 10 gewesen sind. Diese Zahl der Jahre kann ich addieren und subtrahieren; es macht für den Zins als solchen, abgesehen von den Kosten, keinen Unterschied, ob ein Kapital zwei Jahre und dann drei Jahre oder einmal 5 Jahre auf Zinseszins liegt. Man muß also fragen: Wie oft ist die erste Zahl um einen festen Prozentsatz, z. B. 1%, zu vermehren, also mit 101 zu multiplizieren bezw. 101 zu dividieren, um die folgende zu erhalten.

Sind die Zahlen a und b, die Einheit der Zunahme q, so ist die Zahl x die angibt, wie oft a mit q zu multiplizieren ist, um b zu erhalten, der Logarithmus von $\frac{b}{a}$ für die Basis q. Dies ergibt sich auch aus der Gleichung:

$$b = a \cdot q^x$$

daraus folgt:

$$\log b = \log a + x \log q$$
$$x \log q = \log b - \log a$$
$$x = \frac{\log b - \log a}{\log q}$$
$$x = {}^{q}\log \frac{b}{a}$$

Bei der vorletzten Gleichung ist die Basis beliebig und kann daher eine Logarithmentafel mit Basis 10 angewandt werden.

Als Einheit der Ordinaten ist als log. q zu nehmen und jeder Quotient der aufeinanderfolgenden Glieder durch seinen Logarithmus in dieser Einheit auszudrücken. Setzt man so log. q = 1, so ist x = log. a — log. b = log. $\frac{b}{a}$. Summiert man die Logarithmen der Quotienten zu dem Logarithmus des ersten Gliedes der Reihe log. a, so erhält man die Logarithmen der Reihe. Leichter ist es natürlich, zuerst die Logarithmen der Reihe aufzusuchen und von diesen die Logarithmen der Quotienten abzuleiten.

Auch ohne diese Ableitung läßt sich aus der Steilheit oder Flachheit der Kurve der Logarithmen die jeweilige relative Zu- oder Abnahme ersehen. Zugleich wird die absolut Größe und die absolute Zu- und Abnahme der Zahlen selbst wenigstens angedeutet, und zwar kommt die Kurve der Logarithmen derjenigen der ursprünglichen Zahlen um so näher, je kleiner die relativen Unterschiede der letzteren sind. Allerdings muß man es deutlich angeben, wenn die Logarithmen dargestellt sind, um eine Verwechselung zu vermeiden, denn die logarithmische Darstellung kann die Wiedergabe der ursprünglichen Zahlen nicht ersetzen. Es wäre aber unbillig, ihr daraus einen Vorwurf zu machen. Was sie allein ersetzen kann und soll, sind die Kurven der relativen Ab- und Zunahmen.

Die Zuwachsprozente werden nicht nebeneinander gestellt, sondern ihre Logarithmen addiert und so die Vorstellung der Zunahme genau und anschaulich hervorgerufen.

Wenn die prozentuale Zunahme sich gleich bleibt, entsteht eine gerade Linie, und zwar sowohl bei der Kurve der Zuwachsprozente, als bei derjenigen der Logarithmen der Zahlen. Nur ist im letzteren Falle diese gerade Linie schräg gerichtet und drückt durch ihre Neigung die Intensität der Zunahme aus, ebenso, wie die Kurve der prozentualen Zunahme durch ihren Abstand von der 0-Linie.

Nimmt die Zunahme selbst zu, so wenden sich beide Kurven nach oben, und zwar die Kurve der Logarithmen in einem sich weiter krümmenden Bogen, die Kurve der prozentualen Zunahme geht in eine schräge, gerade Linie über, wenn die Zunahme der Zunahme konstant ist.

Summiert man die Differenzen der Logarithmen nicht wieder zu der Kurve der Logarithmen der Reihe, sondern stellt sie auf die Abszissenachse nebeneinander, so hat man die Logarithmen der prozentualen Zunahme, und kann an diesen die relativen Unterschiede der relativen Zu- und Abnahme beobachten.

Die Logarithmenkurve hat ferner den Vorteil, daß der untere leere Teil des Diagramms fortgelassen werden kann, da man ohne Fehler die Einheit doppelt, 10-, 100- oder 1000fach größer wählen kann. Eine Multiplikation wird durch einfache vertikale Verschiebung ausgeführt, ohne die Form der Kurve zu ändern. Die Tendenzen der Entwicklung treten klar hervor. Wenn sich eine Größe aus sich heraus vermehrt, ergibt sich bei gleichbleibender relativer Zunahme eine geometrische Progression. Eine solche erscheint bei der Kurve der Zahlen als eine konkave, immer steiler ansteigende Kurve und erweckt so die falsche Vorstellung, als ob die Tendenz der Zunahme selbst noch im Steigen begriffen wäre. Die Kurve der Logarithmen vermeidet dies und läßt erkennen, wie bei vielleicht noch steigender absoluter Zunahme diese relativ sinkt.

So läßt sich auch sofort feststellen, wann die Entwicklung gehemmt und gefördert worden ist und welche Richtung sie nimmt, welche Resultate bei Fortbestand der gegenwärtigen Entwicklungstendenzen zu erwarten sind.

Falsch wäre es allerdings, logarithmische Kurve da anzuwenden, wo man es nicht mit einem Phaenomen zu tun hat, das sich aus sich selbst heraus entwickelt, oder das einem solchen proportional ist, sondern mit einem solchen, das den Grund seiner Hemmung durch sein Wachstum schafft und vermehrt.

Anhang

1. Graphische Ausgleichung zu einer Reihe von Parabeln

Will man die gebrochene Linie durch Ausgleichung in eine krumme verwandeln, so kann man dies graphisch leicht in folgender Weise erreichen:

Man teilt bei gleichem Abstand der Ordinaten die Verbindungsstrecken zwischen je zweien in vier gleiche Teile und verbindet die jedem Eck benachbarten Teilpunkte. Dann sind die zugefügten und fortgenommenen Dreiecke in ihrer Summe gleich, wenn man vor der ersten und nach der letzten Ordinate horizontale oder sonst parallele Linien gezogen hat. Sind die Höhen der Ordinaten a, b, c, so sind die Dreiecke, wenn der Abstand der Ordinaten gleich i ist,

$$\frac{i}{4}\left[\left(\frac{\frac{3a+x}{4}+\frac{3a+b}{4}}{2}-a\right)+\left(\frac{\frac{3b+a}{4}+\frac{3b+c}{4}}{2}-b\right)+\left(\frac{\frac{b+3c}{4}+\frac{3c-x}{4}}{2}-c\right)\right]$$

$$=\frac{i}{4}\left(\frac{x-2a+b}{8}+\frac{a-2b+c}{8}+\frac{b-c+x}{8}\right)=0$$

Die Summe der Dreiecke, die positiv werden, ist gleich derer der Negativen. x bedeutet den Unterschied zwischen der ersten und letzten Ordinate und der gedachten nächsten.

Wiederholt man dieses Verfahren, so nähert man sich stets mehr und mehr einer krummen Linie, die man dann mit der Hand mit genügender Genauigkeit einzeichnen kann. Diese Linie besteht aus Parabeln, die sich zwischen den von der Ausgleichung nicht mehr verschobenen Punkten für $\frac{a+b}{2}$, $\frac{b+c}{2}$ usw. erstrecken, d. h. denjenigen Punkten, deren Abszissen und Ordinaten die Mitte zwischen denen für a und b, b und c bilden, auf der Figur die Mittelpunkte der ersten Verbindungslinien. (Siehe Fig. 3.)

Daß die Linien Parabeln sind, kann man daraus ersehen, wenn man die Durchschnittsbildung nach nebenstehendem Schema durchrechnet und die zweiten Differenzen bildet. Die Durchschnittsbildung kann man sich beliebig weit fortgesetzt denken, die zweiten Differenzen werden stets zwischen den Grenzen Konstante bilden; die Figur nähert sich also je einer Parabel für jedes Stück, dessen Grenzen die Mittelpunkte der ersten Verbindungslinien bilden. Jedes dieser Stücke ist nur von drei ursprünglichen Zahlwerten abhängig; da die einmal festen Punkte ihre Lage nie mehr ändern, braucht man die jenseits liegenden Werte nicht mehr zu beachten. Einer solchen Ausgleichung der Entwicklung einer Bestandsmasse entspricht die Ausgleichung der Treppenfigur der Bewegungsmasse durch Verbindung der Mittelpunkte der Zeitstrecken.

2. Graphische Interpolation durch Parabeln

Verschieden von der Ausgleichung ist die Aufgabe der Interpolation. Es sind nicht die zufälligen Fehler von Daten zu beseitigen oder aus der oscillierenden beobachteten Reihe eine stetige Entwicklung zu entnehmen. Die gegebenen Daten sollen vielmehr als feststehend behandelt und nur ergänzt werden. Es ist also eine Linie zu zeichnen, die die gegebenen Ordinaten hat, und sie in einer Weise verbindet, die dem vermuteten Entwicklungsgesetz entspricht.

Die einfachste Form ist gradlinige Verbindung; sie hat den Mangel, daß die Maxima und Minima und in geringerem Maße die anderen Punkte plötzliche Änderungen der Entwicklungsrichtung darstellen.

Um einer allmählich ihre Richtung ändernden Kurve näher zu kommen, kann man annehmen, daß die Richtung derselben an jedem gegebenen Punkt der Verbindung zwischen den benachbarten Punkten parallel ist. Zeichnet man diese Parallelen bis zu $^1/_4$ des Abstandes der gegebenen Ordinaten, verbindet die Endpunkte, so hat man eine gebrochene Linie, die man in derselben Weise zu einer Kurve umgestalten kann, wie wenn sie aus einer Treppenfigur entstanden wären. Man gewinnt so eine fortlaufende, aus Parabeln bestehende Linie, die die empirisch gegebenen Punkte verbindet. (Siehe Fig. 3.)

Die Treppenfigur der Bestandsmasse ist dann so umzugestalten, daß die Mittelpunkte der Differenzen der Ordinaten und die um den Abstand von deren Verbindungslinie nach der entgegengesetzten Richtung verschobenen Mittelpunkte der wagrechten Strecke verbunden werden. Die Fläche für jede beobachtete Zeitstrecke bleibt dann unverändert.

3. Vergleichung von Diagrammen durch drehbare Ausschnitte

Die Vergleichung von Zahlenreihen wird also zunächst durch die graphische Darstellung der ursprünglichen Zahlen selbst erleichtert, sodann durch Darstellung abgeleiteter Reihen, die die Aufgabe haben, einen Teil der Unterschiede zu beseitigen, um einen andern um so deutlicher hervortreten zu lassen.

Oft wird man nicht auf den ersten Blick erkennen können, wie die Reihen am besten zur Vergleichung zu bringen sind, z. B. wenn die Veränderungen der einen Reihe in umgekehrter Richtung erfolgen, als die der anderen, wenn sie ihnen zeitlich nachfolgen, wenn sie nicht den Unterschieden der ganzen Reihe, sondern nur den Differenzen entsprechen. Für diese Fälle wäre es erwünscht, durch vorläufige Versuche die beste Art der Vergleichung ausfindig zu machen.

Soweit es sich dabei nur um Addition, Subtraktion, Multiplikation und Division handelt, kann ein einfacher Kunstgriff hierzu behilflich sein. Man zeichnet nämlich Diagramme zweier Zahlenreihen, in der bekannten Form der Verbindungslinien der Ordinaten eines Koordinatensystems. Das eine der entstehenden Polygone schneidet man aus und hält es vor das Papier, auf dem das andere sich befindet, so kann man durch Drehungen des Polygons sehen, ob durch Anwendung der vier einfachsten Rechnungsarten eine größere Ähnlichkeit erzielt werden kann. Dreht man nämlich das ausgeschnittene Diagramm um eine von rechts nach links verlaufende Achse, indem man die obere Hälfte hebt, die untere senkt, so verkürzt sich perspektivisch die Höhe der dargestellten ausgeschnittenen Figur, während die darunter liegende unverändert bleibt. Der Erfolg ist derselbe, als wenn man die Zahlen des beweglichen Diagramms durch einen konstanten Faktor dividiert hätte. Blickt man schräg auf die daliegende Figur und senkrecht auf die ausgeschnittene bewegliche, so ist der Erfolg umgekehrt, die Größen der liegenden Figur werden durch einen kon-

stanten Divisor dividiert. Läßt man das Licht schräg auf die daliegende Figur von einer gegenüber oder hinter dem Beschauer befindlichen Lichtquelle fallen, und hält die ausgeschnittene Figur davor, so daß sie einen Schatten auf die liegende wirft, so kann man durch Drehung die Grenze des Schattens dem Diagramm soweit als möglich nähern. Bewegen sich die beiden Kurven in entgegengesetzter Richtung, so muß die bewegliche Figur um ihre Achse gedreht werden, bis die Rückseite oben erscheint.

Ebenso wie durch Drehung der Figuren Division und Multiplikation der zugrundeliegenden Zahlenreihe durch jede beliebige Größe sofort veranschaulicht werden, ebenso wird durch Verschiebung der beweglichen Figur nach oben der unten (d. h. ▲ bezw. ▼) die Addition und Subtraktion einer konstanten Größe zu oder von allen Gliedern der Zahlenreihe dargestellt. Am genauesten läßt sich die Vergleichung auf diese Weise erzielen, wenn man den Schatten des Sonnenlichtes entstehen läßt, weil deren Strahlen so gut wie parallel sind. Hat man jedoch einige Übung darin, das Bild einer Figur im Gedächtnis zu behalten, es genügt, sie schräg zu betrachten, um festzustellen, ob ihre relative Zu- und Abnahme übereinstimmt oder nicht.

Hat man auf diese Weise gefunden, an welchen Punkten die Ordinaten der beiden Figuren durch Drehung zusammenfallen, so kann man leicht berechnen, durch welchen Divisor die Zahlen der einen Reihe dividiert werden müssen, und welche Größe zu addieren oder subtrahieren ist, um die größtmöglichste Übereinstimmung zu erzielen. Man hat zu diesem Zweck die als gleich erscheinenden Strecken beider Figuren zu messen, und das Verhältnis zu berechnen.

Die Drehung kann übrigens nicht nur um die von links nach rechts verlaufende Achse, die Verschiebung nicht nur in der Richtung der senkrecht dazu stehenden y-Achse geschehen. Wenn nämlich die Phänomene der einen Reihe schneller verlaufen als die der anderen, so kann dies durch eine Drehung um die y-Achse ausgeglichen werden, wenn sie denen der anderen Reihe seitlich nachfolgen, so ist eine Verschiebung nach rechts geeignet, den Zusammenhang aufzudecken.

Soll eine solche Beziehung der beiden Kurven untersucht werden, daß die x- und y-Achse vertauscht werden, so kann eine Drehung um eine Linie angewandt werden, die im Winkel von 45% zwischen der x- und y-Achse nach rechts oben ansteigt. Ist z. B. für die eine Kurve

$$x_1 = 1;\ x_2 = 2;\ x_3 = 3;\ x_4 = 4$$
$$y_1 = 2;\ y_2 = 4;\ y_3 = 8;\ y_4 = 16$$

für die andere umgekehrt

$$x_1 = 2;\ x_2 = 4;\ x_3 = 8;\ x_4 = 16$$
$$y_1 = 1;\ y_2 = 2;\ y_3 = 3;\ y_4 = 4$$

so ist für die Betrachtung der Kurven in der üblichen Lage eine Übereinstimmung nicht zu vermuten. Dreht man dagegen die eine Figur um die im Winkel von 45° verlaufende Linie, so daß sich ihre x-Achse mit der y-Achse der anderen Figur deckt, so fällt die vollkommene Übereinstimmung in die Augen.

Diese Methode kann ebenfalls wieder mit den Drehungen und Verschiebungen in anderer Richtung verbunden werden.

Hat man auf diese experimentelle Methode die einfachen Rechnungsoperationen gefunden, sowie festgestellt, ob eine Kurve durch Vertauschung der x- und y-Achse der andern ähnlich wird, so kann zunächst die Reihen der Zahlen dementsprechend umgestalten, und sie daraufhin untersuchen, wie groß die Korrelation des Abstandes der einzelnen Glieder vom Mittel, sowie die Korrelation der Zunahme oder Abnahme zwischen entsprechenden Gliedern ist.

4. Kontinuität

Eine Verallgemeinerung dieser Gedanken ergibt folgendes:

Eine Verbindung der durch beobachtete Zahlen im Koordiantensystem bestimmten Punkte kann die Vorstellung erwecken,

1. daß diese Punkte der Ausdruck der Wirklichkeit sein können;
2. daß sie es tatsächlich sind, und durch Beobachtung oder Schlußfolgerung ihre Lage mit hinreichender Genauigkeit bestimmt wurde;
3. daß die dargestellten Punkte gleich bedeutungsvoll sind.

Die erste dieser Vorstellungen muß bei jeder Verbindungslinie von solchen Punkten, den Endpunkten der Ordination, eintreten. Die zweite und dritte kann dadurch vermieden werden, daß die Verbindungslinien die Form eines Polygons annehmen, dessen Ecken die beobachteten oder die von gleicher Bedeutung angenommenen Punkte deutlich hervortreten lassen, oder daß diese Punkte besonders bezeichnet werden (z. B. durch Sternchen) oder endlich, daß die zu diesen Punkten führenden Ordinaten durch Linien oder schmale Rechtecke bezeichnet werden.

Diese erste Vorstellung, daß die Punkte der Verbindungslinien möglich sind, setzt voraus, daß beide dargestellten Erscheinungen zwischen den beobachteten Punkten dauernd, oder nur mit kleinen, graphisch nicht darstellbaren Unterbrechungen bestehen, daß sie in Beziehung zu einander stehen und ihre Größe nur allmählich ändern. Ob dies der Fall ist, kann das Urmaterial der Statistik oder sonstige Erfahrungen lehren.

Solche dauernden, allmählich sich verändernden Größen sind:

1. die Zeit, d. h. die Summe der abgelaufenen Zeit;
2. die Reihe der möglichen Größen irgendwelcher Art, mit Ausnahme derjenigen, die nur in ganzen Zahlen denkbar sind, deren Unterschiede zu groß sind, um graphisch vernachlässigt zu werden.

Diese beiden Arten von Größen ändern sich nicht nur allmählich sondern auch gleichmäßig, daher hat jeder Punkt ihrer Achse (in der Regel wird für sie die Abszissenachse gewählt) gleiche Bedeutung. Sie sind die „unabhängige Veränderliche".

Allmählich, aber nicht notwendig gleichmäßig, können sich ändern:

3. Bestandsmassen;
4. Bewegungsmassen in ihrer Beziehung zu den Zeiträumen, in denen sie verlaufen. Für die einzelnen Zeitpunkte sind sie entweder unendlich klein oder sie verlaufen nicht kontinuierlich.
5. Andere Beziehungszahlen zwischen allmählich sich ändernden Größen;
6. die Merkmale, d. i. meßbare Eigenschaften von Objekten, die in der Reihenfolge ihrer Größe geordnet sind, und
7. die Zahlen der Objekte, bei denen diese Merkmale vertreten sind. Als Merkmale kann man auch die Größen verschiedener Bestands- oder Bewegungsmassen auffassen, die durch ihre Gleichzeitigkeit oder unter einem anderen Gesichtspunkt zusammengefaßt sind.

Eine Verbindung der Punkte im Koordinatensystem ist nicht zulässig, wenn die Kontinuität so erheblich unterbrochen ist, daß ein falsches Bild entstehen würde. Eine vollkommene Kontinuität ist nicht erforderlich; genau genommen liegt diese nach den Ergebnissen der Naturwissenschaft ja auch bei Erscheinungen nicht vor, die uns so ununterbrochen erscheinen, wie das fließende Wasser und das ruhige Licht, bei denen aber theoretisch Atome und Lichtwellen unterschieden werden können. Die Statistik betrachtet die Dinge von

weitem und läßt viele Einzelheiten zu einem Gesamtbild verschmelzen, wie die Blätter eines Baumes und die Bäume eines Waldes.

Die Kontinuität ist also ein relativer Begriff, und es ist möglich, Erscheinungen in einer Form so darzustellen, daß die tatsächlichen Unterbrechungen den Anschein der Kontinuität nicht mehr stören.

Ein Mittel hierzu ist die Summierung der Einzelheiten. Sie sind dann im Verhältnis zur Gesamtheit verschwindend klein; sie zeigen sich nur in dem Wachsen der Summe. Die Bildung der Bestandsmasse aus der Bewegungsmasse ist die Summierung positiver und negativer Einzelvorgänge in der Zeit; dasselbe Prinzip kann aber auch in andern Fällen angewandt werden, so bei Kumulierung der Zahlen der nach ihrem Merkmal geordneten Objekte.

Ein zweites Mittel ist das, die unterbrochene Größe mit einer kontinuierlichen in Beziehung zu setzen. So treten die Einzelvorgänge einer Bewegungsmasse in einem Augenblick auf und verschwinden wieder; die Beziehung ihrer Zahl zu einer sie umschließenden Zeitstrecke oder einem Raum ändert sich nur unmerklich, wenn dieser groß genug gewählt. Indem man jeden Zeitpunkt als Mittelpunkt einer Zeitstrecke auffaßt, die z. B. ein halbes Jahr rückwärts und ein halbes Jahr vorwärts umfaßt, kann man eine kontinuierliche Reihe bilden, und die einzelnen beobachteten Punkte durch Linien verbinden. In ähnlicher Weise kann man andere Beziehungen graphisch darstellen. Z. B. die Größenmaße der Kinder als Funktion zur Größe der Eltern, das Alter der heiratenden Männer zu dem der Frauen und umgekehrt. Hierbei ist von der zweiten der abhängigen Größen der Durchschnitt zu bilden, damit eine einheitliche Linie entsteht; eventuell können auch die Maxima und Minima und die Quartilen gegeben werden.

Ist es nicht möglich, in dieser Weise beide Größen zu einander in kontinuierliche Beziehungen zu bringen, so würde eine Verbindungslinie der Punkte irreführend sein, und sind nur die unverbundene Ordinaten oder ihre Endpunkte darzustellen. Die Ordinaten lassen die Größe anschaulicher erscheinen; sind sie aber sehr zahlreich, so werden sie unübersichtlich. Dann ist ein Punkteschwarm in Koordinatensystem das beste Ausdrucksmittel der Zusammengehörigkeit je einer Größe der Ordinate zu einer der Abszisse. Gibt man den Punkten verschiedene Farben, so kann man sie nach einem weiteren Gesichtspunkt differenzieren. Für jeden Punkteschwarm kann man graphisch nach dem Augenmaß oder durch Berechnung eine Linie als Achse zeichnen, deren Neigung zwischen den Neigungen der Linien liegt, die die Punkte der Durchschnitte der einen Erscheinung als Funktion der anderen und der zweiten als Funktion der ersten verbinden.



B. Merkmals- und Häufigkeitskurven.

Prüfung statistischer Hypothesen durch graphische Darstellung

Die genaue und vollständige Wiedergabe eines Gegenstandes ist der Wissenschaft ebenso unmöglich wie der Kunst. Die künstlerische und wissenschaftliche Beschreibung und Erfassung muß sich darauf beschränken, das für den besonderen Zweck Wesentliche hervorzuheben und in der Form darzustellen, die sich dem Objekt und zugleich dem Ausdrucksmittel am besten anpaßt. Damit ist zugleich gesagt, daß der schöpferischen Phantasie des Beschauers, Lesers oder Hörers ein Spielraum bleiben wird, den er mit mehr oder weniger klaren und richtigen Vorstellungen ausfüllen kann und ausfüllen muß, wenn er aus der Schilderung sich ein lebendiges Bild der Wirklichkeit machen will.

Die beste Darstellung kann diese ergänzende Phantasie nicht entbehrlich machen, sie kann nur versuchen, sie zu unterstützen und auf den richtigen Weg zu leiten.

Dies gilt in erhöhtem Maße von einer statistischen Wiedergabe der Wirklichkeit. Die Statistik kann die Dinge zu einzelnen Zeitpunkten messen und zählen und so Anhaltspunkte gewinnen für die Entwicklung, die in der ununterbrochen fließenden Zeit verläuft; die unendlich zahlreichen und unendlich feinen denkbaren Messungen, die einen Gegenstand vollständig beschreiben würden, müssen auf Messungen weniger Merkmale mit oft nur annähernder Genauigkeit beschränkt werden, um zahlenmäßig wiedergegeben werden zu können.

Diese Beschränkung liegt nicht nur im Interesse des Schreibers, sondern auch des Lesers, sie ist nicht nur die Voraussetzung der Herstellbarkeit einer Statistik, sondern noch mehr ihrer Verständlichkeit. Die Menge der Zahlen läßt sich durch Aufwand von mehr Mühe und Kosten vermehren; der Aufnahmefähigkeit des Lesers sind aber Grenzen gezogen, vor allem der Fähigkeit, mehrere Zahlen gleichzeitig zu erfassen, zu vergleichen und in ihren gegenseitigen Beziehungen zu würdigen.

Vorteile der graphischen Darstellung

Eine sehr viel größere Menge von Größenvorstellungen als durch Zahlen kann dem Beschauer durch graphische Darstellungen, insbesondere „Kurven", übermittelt werden. Eine einzige Kurve in einem Koordinatensystem enthält unendlich viele Punkte, die dem Beschauer alle gleichzeitig zum Bewußtsein kommen und Größen bedeuten; zugleich gibt sie die Vorstellung ihrer gegenseitigen Beziehung durch den Vergleich aller denkbaren Ordinaten und Abszissen, die Reihenfolge der Größen wird durch das Ansteigen der Kurve ausgedrückt, jede kleinste Wendung, jede Eigenart der Kurve hat ene Bedeutung für die dargestellten Größen. Hieraus folgt zweierlei: Erstens, wenn es darauf ankommt, beim Beschauer und beim Verfasser selsbt eine möglichst große Zahl von Zahlen und Zahlenbeziehungen gleichzeitig zum Bewußtsein zu bringen, ist die Kurve den Ziffern überlegen; zweitens geht ihr Reichtum an Ausdrucksmöglichkeiten noch über das Maß des durch Zählung und Messung gewinnbaren Materials hinaus.

Die Folge ist, daß in vielen Fällen die Darstellung Ausdrucksmittel enthält, die nicht verwertet werden und durch ihre Anwesenheit stören oder gar zu falschen Vorstellungen Anlaß geben können. Dies gilt insbesondere von Kurven im Koordinatensystem und Figuren, die in einem Koordinatensystem eingetragen sind oder eine zeitliche oder sachlich quantitative Reihe bilden. Bei den Kurven sind nur einzelne Punkte der Beobachtung entnommen, die Verbindungslinien durch Schätzung gewonnen; vermeidet die Zeichnung die Verbindungslinien und beschränkt sich auf einzelne Ordinaten oder Rechtecke, so gibt sie dem Auge Gelegenheit, die Kontinuität durch eine gedachte Linie herzustellen.

Besagen z. B. die Zahlen der Tabelle, daß eine Bevölkerung im Jahre

1900 3720
1905 4123
1910 4453

betragen habe, und stellt man dies graphisch dar., so kann man drei Linien als Ordinaten im gleichen Abstand hinstellen. Der Beschauer weiß, daß die Bevölkerung in der Zwischenzeit vorhanden war und vermutet, daß sie kontinuierlich zugenommen hat. Als Abbild dieser Kontinuität erscheint ihm die Verbindungslinie der Endpunkte der Ordinaten und ist eine solche nicht gezogen, so stellt er sie vor. Dasselbe gilt von Darstellungen einer Bewegung als solcher, zum Beispiel der absoluten oder relativen Zunahme.

Die obigen Zahlen ergeben im ersten Jahrfünft eine Zunahme von 403, im zweiten von 344; der Beschauer wird die Ordinaten in Gedanken verbinden, auch wenn sie unverbunden aufgestellt sind, oder die Zunahme durch Rechtecke ausgedrückt ist, deren Grundlinie der Zeitstrecke, deren Flächeninhalt dem Betrag der Zunahme proportional sind. Diese Vorstellung vervollständigt die Daten der Statistik durch Hypothesen, und sucht sich auf diese Weise der Wirklichkeit zu nähern, sie ist aber von dem Beschauer abhängig und unkontrollierbar. Es kann sein, daß sie zu Ergebnissen führt, die untereinander und mit anderen feststehenden Tatsachen im Widerspruch stehen. Dies kann der Zeichner verhindern, wenn er neben oder anstelle der Figur aus den Daten selbst eine korrekte Vervollständigung zieht, die die Kontiunität herstellt, ohne mit der Wirklichkeit in Widerspruch zu geraten.

Wertvoll ist es, dabei auch die wirklich statistisch erfaßten Zahlen deutlich hervortreten zu lassen und von den hypothetischen Verbindungen zu trennen, damit der Beschauer die Hypothese nachprüfen und von den sicheren Ergebnissen unterscheiden kann und jeder Anschein tendenziöser Entstellung vermieden wird.

Die genaue Wiedergabe der Zahlen und die einwandfreie Konstruktion derjenigen Ergänzungen, die sich aus der Natur der Dinge und anderen Daten ergeben, ist besonders notwendig, wenn die Kurven zur Grundlage weiterer Schlußfolgerungen gemacht werden sollen. Zu diesem Zweck ist es unter Umständen notwendig, eine Ausgleichung vorzunehmen und die Haupttendenz der Kurve von ihren kleinen Schwankungen zu trennen.

Im Folgenden soll gezeigt werden, daß es mit Hilfe einfachster geometrischer Konstruktionen möglich ist, die Ausgleichung so durchzuführen, daß der Durchschnittswert der Ordinaten und Abszissen unverändert bleibt; und zwar soll die Anwendung dieser Methode auf die Untersuchung sachlich quantitativer Reihen[1]) erörtert werden; d. h. solcher, die die Messungen eines Merkmals bei einer Reihe gleichartiger Gegenstände enthalten.

[1]) Zizek, Mittelwerte.

Solche Reihen kann man auf zwei Weisen graphisch darstellen: entweder kann man das Hauptgewicht auf die Messungen des Merkmals legen, und diese durch Ordinaten ausdrücken, die in gleichen Abständen nach ihrer Größe geordnet sind, wie eine Reihe von Soldaten, oder man kann die Häufigkeit darstellen, d. h. die Zahl der Objekte, die auf gleiche Abstufungen entfällt. Im ersten Fall erhält man die Kurve der Grade des betreffenden Merkmales, im letzteren die Häufigkeitskurve.

Beide stehen in einem engen Zusammenhang. Die Abszissen der Merkmalskurve geben nämlich an, das wievielste Glied der Reihe durch die Ordinaten dargestellt sind, also zugleich, wie groß die Summe der Objekte bis zu diesem Grad des Merkmals ist; dreht man sie daher im Sinne der Uhrzeiger um einen rechten Winkel, so muß sie bei richtiger Konstruktion mit einer Kurve zusammenfallen, deren Ordinaten die Zahl der gemessenen Objekte, deren Abszissen die Grade des Merkmals darstellen. Dabei ergänzen sich die von den Kurven und ihren Achsen eingeschlossenen Flächen zu einem Rechteck. Die Gruppen gleicher Objekte werden also zu einander addiert, daher wird diese Kurve, die von Galton für Körpermessungen und von Bowley für Lohnstatistik angewandt wurde, „Kumulative Kurve" genannt.

Während die Zahl der Objekte bei diesen Kurven durch Längen dargestellt wird, und zwar die Gesamtzahl durch die größte Ordinate bezw. Abszisse, gibt die gewöhnliche Häufigkeitskurve die Gesamtzahl durch die von ihr eingeschlossene Fläche, die auf ein Intervall des Merkmals entfallende Zahl durch den Teil der Fläche, der auf den Ordinaten für die Grenzen des Interwalls und den zwischen ihnen liegenden Teil der Kurve eingeschlossen ist.

Bei genauer Darstellung muß also die von der kumulativen Kurve erreichte Höhe (bezw. die von der Kurve der Merkmale erreichte Breite) der von der Häufigkeitskurve eingeschlossenen Fläche proportional sein; bezeichnet man als Einheit der Fläche das von der Ordinate und der Abszisse für 1 gebildete Rechteck, so muß die Häufigkeitskurve diese Fläche ebenso oft einschließen, als die größte Ordinate der kumulativen Kurve Einheiten enthält; bei ungenauer Konstruktion ergeben sich Abweichungen; auch fällt dann eventuell die kumulative Kurve nicht mit der Kurve der Merkmale zusammen.

Soll die Konstruktion der Kurven den dargestellten Zahlen genau ent-
Die benutzte Tabelle kann nämlich entweder bestimmte Werte für die Messungen angeben, oder nur die Grenzen, zwischen denen die Messungen liegen. Im ersteren Fall könnte man die Zahl der Objekte bei denen die Messung zutrifft, mit „Messungsgruppe", im letzteren mit „Intervallgruppe" bezeichnen. In beiden Fällen können die Intervalle gleichmäßig oder ungleichmäßig abgestuft sein. So ergeben sich vier Möglichkeiten für jede der beiden Kurvenformen:

„Messungsgruppen" oder „Intervallgruppen"; beide mit gleichen oder ungleichen Intervallen.

I. Kumulative und Merkmalskurve

1. Für gleiche Intervalle

Hat die Tabelle eine solche Form, daß die einzelnen bestimmten Messungen und die dazu gehörigen Zahlen der Objekte angegeben sind, so ist die genaue Form der Merkmals- sowie der kumulativen Kurve treppenförmig. Die Merkmalskurve beginnt bei der Abszisse 0 mit der Ordinate der höchsten Eigenschaft;

sie behält diese Ordinate bei, bis die Abszisse der Zahl der Objekte dieser Eigenschaft erreicht ist; dann sinkt die Ordinate plötzlich auf die dem nächsten Grade des Merkmals entsprechende Größe, bis die Abszisse den Wert für die Zahl der Objekte in beiden Gruppen zusammen entspricht usw.

Die kumulative Kurve hat umgekehrt zunächst die Ordinate 0 bis die Abszisse des niedrigsten Merkmals erreicht ist, steigt dann plötzlich um die Zahl der Objekte der niedrigsten Gruppe, um bis zur Abszisse der nächsten gemessenen Eigenschaft wieder wagrecht zu verlaufen. Beide Treppenlinien fallen durch Drehung genau zusammen und diejenige, der Merkmale stimmt mit einer Reihe von Rechtecken überein, deren Höhen, deren Merkmale und deren Breiten der Zahl der Objekte jeder Gruppe proportional ist. Die kumulative Kurve ergänzt diese Figur zu einem Rechteck, dessen eine Seite der Gesamtzahl der Objekte, dessen andere Seite dem höchsten Grade des gemessenen Merkmals proportional ist.

Die Darstellung der Messungen durch eine treppenförmige Reihe von Rechtecken entspricht aber nur dann der Wirklichkeit, wenn eine wirklich genaue Messung möglich ist.

Gehen dagegen die Grade des zu messenden Merkmals in unmerklichen Abstufungen ineinander über, wie die Längenmaße von Körpern, die Zeit oder andere kontinuierliche Größen, so sind auch die anscheinend genauesten Messungen nur schätzungsweise Annäherungen an den nächsten derjenigen Werte, die in die Tabelle aufgenommen werden sollen. Sie sind eine ungenaue Ausdrucksweise für eine Intervallgruppe; z. B. wenn Längemaße in cm angegeben werden, bedeutet genau genommen

170 = 169,5 bis 170,5
171 = 170,5 bis 171,5
172 = 171,5 bis 172,5 usw.

In derartigen Fällen entspricht es dem Wesen der beobachteten Zahlen, sie nicht als Messungsgruppen, sondern als Intervallgruppen aufzufassen, deren Grenzen um ein halbes Normalintervall von der angegebenen Zahl entfernt sind.

Darüber, wie die wirklichen Maße in diesem Intervall verteilt sind, geben die einzelnen Messungen keine Anhaltspunkte. Es ist also zunächst die Wahrscheinlichkeit für jeden der unendlich vielen Grade gleich groß. Dieser Auffassung entspricht eine gradlinige schräge Verbindung für die an den Grenzen der Intervalle geltenden Punkte. Die unendlich viele Punkte enthaltende Linie ist ein getreuer Ausdruck der unendlich vielen Möglichkeiten, die gleichmäßige Veränderung der Abszissen und Ordinaten stellt die als gleich angenommene Verteilung der Möglichkeiten dar. Man teilt also bei der Merkmalskurve jedes der gleichen Intervalle zwischen den Messungen in zwei Hälften und verbindet die Mittelpunkte. Ist eine Messung nicht vertreten, so läßt man in der Merkmalskurve eine senkrechte, in der kumulativen Kurve eine wagrechte Stelle entstehen.

2. Für ungleiche Intervalle

Sind die Messungen angegeben und haben diese ungleiche Intervalle, so ist wieder die treppenförmige Figur die richtige Wiedergabe der Wirklichkeit, wenn anzunehmen ist, daß die Messungen genau sind. Sind diese dagegen nur der ungenaue Ausdruck einer Reihe von möglichen Werten des zu messenden Merkmals, so ist wieder die treppenförmige Figur durch Hinzufügung und Hinwegnahme von Dreiecken in ein Polygon zu verwandeln. Die Dreiecke sollen der zu erwartenden Ungenauigkeit entsprechen.

Ist anzunehmen, daß die Ungenauigkeit in allen Teilen der Reihe gleich groß ist, und mit gleicher Wahrscheinlichkeit nach beiden Richtungen stattgefunden kaben kann, so ist wieder ein jedes Rechteck der Merkmalskurve durch ein Trapez zu ersetzen, dessen eine Seite um den Betrag der Ungenauigkeit erhöht, dessen andere Seite um diesen Betrag erniedrigt ist. Da infolge der Ungleichheit der Intervalle die Seiten dieser Trapeze nicht zusammenfallen, entsteht eine Figur, die aus schrägen abwärts führenden Linien besteht, die durch teils abwärts, teils aufwärtsführende senkrechte Strecken verbunden sind. Die durch die Ordinaten dargestellten Messungen sind also nicht mehr durchweg nach der Größe geordnet. Stellt man diese Ordnung wieder her, so muß man die Abszissen der Trapeze für jeden Punkt der Ordinatenachse addieren, was man dadurch ausführen kann, daß man die Abszissen für jeden Eckpunkt addiert, und die gewonnenen Punkte verbindet. Man hat so die aufwärts gerichteten rechtwinkeligen Dreiecke, um die die Trapeze das bisher erreichte Minimum übersteigen, in gleiche Dreiecke verwandelt, die mit der Spitze an das vorhergehende Trapez herangezogen sind. So ist die Kontinuität hergestellt, ohne die Fläche, und somit ohne die Gesamtsumme aller Messungen und deren Durchschnitt zu verändern.

Ist dagegen anzunehmen, daß die Genauigkeit der Messungen um so größer ist, je näher diese einander kommen, so ist das Intervall zwischen je zwei Messungen zu halbieren. Ist dabei weiter anzunehmen, daß die Objekte jeder Gruppe mit gleicher Wahrscheinlichkeit zu groß oder zu klein gemessen sind, so ist auch die Zahl der Objekte jeder Gruppe zu halbieren. Es sind also alle wagrechten und senkrechten Stücke der Treppenfigur zu halbieren und die Mittelpunkte zu verbinden. Aus jedem Rechteck entstehen so für die Merkmalskurve zwei Trapeze. Bei dem ersten und letzten Rechteck ist zu untersuchen, ob eine Abweichung nach beiden Seiten möglich ist, oder ob ein Minimum oder Maximum vorliegt. Im ersteren Fall ist die schräge Linie bis zur Ordinatenachse bezw. bis zur letzten Ordinate festzusetzen, im zweiten ihre Neigung so zu ermäßigen, daß das Maximum bezw. Minimum nicht überschritten wird.

Diese Figur ist leicht zu zeichnen, hat aber den Nachteil, daß die zugrundeliegende Annahme schwer nachzuweisen sein wird, und bei großer Unregelmäßigkeit der Intervalle unwahrscheinlich ist. Außerdem sind die Abweichungen der Trapeze von den Rechtecken nicht symmetrisch, die hinzugefügten Dreiecke können im ganzen größer oder kleiner sein, als die hinweggenommenen. Die Summe und der Durchschnitt aller Messungen weichen also von der aus der Treppenfigur und der Tabelle sich ergebenden ab. Diese Form der Darstellung hat also nur dann Anspruch auf Genauigkeit, wenn die zugrundeliegende Annahme durch wichtige Gründe gestützt werden kann, oder wenn z. B. außer den Einzelmessungen noch eine Gesamtmessung aller Objekte vorliegt, die dem Flächeninhalt des Polygon näher kommt als die Summe der Rechtecke.

Sind endlich die ungleichen Intervalle in der Weise entstanden, daß alle Objekte nach dem Grade eines Merkmals geordnet, in Gruppen geteilt, diese Gruppen im ganzen gemessen und daraus der Durchschnitt berechnet worden ist, so ist keine der bisher angegebenen Darstellungen genau. Weder darf eine zickzackförmige Figur entstehen, weil diese eine unrichtige Reihenfolge voraussetzen würde, noch darf der Durchschnitt der Merkmale zwischen zwei Grenz-Ordinaten, also auch der Flächeninhalt der beiden Trapeze, von dem des Rechtecks abweichen, an dessen Stelle sie getreten sind. Die korrekte Darstellung muß in diesem Falle gleiche Dreiecke von jedem Rechteck hinwegnehmen und hinzufügen. Das wird dadurch erreicht, daß man die Intervalle, also die senkrechten Stücke der treppenförmigen Figur der Merkmale, in gleiche

Stücke teilt, und die die wagrechten Strecken im umgekehrten Verhältnis der angrenzenden Stücke teilt. Zu diesem Zwecke notiert man den Schnittpunkt der Verbindungslinie der Mittelpunkte oder der Endpunkte der angrenzenden senkrechten Stücke auf dem wagrechten Stück. Er teilt dieses im Verhältnis der angrenzenden Strecken. Dann vertauscht man die Teile[1]) und verbindet ihren Grenzpunkt mit den Mittelpunkten der senkrechten Stücke. Das hinzugefügte Dreieck ist dann gleich dem hinweggenommenen, und die Flächeninhalte der beiden Trapeze sind gleich dem des Rechtecks. Mithin ist auch der Durchschnitt der Ordinaten für jede Gruppe und die ganze Figur unverändert.

Aus dem Bisherigen geht hervor, daß die Bezeichnung der Gruppen durch eine einzige Messung nur dann den Tatsachen entspricht, wenn diese Messung wirklich genau sein kann; andernfalls ist sie nur der Ausdruck für ein Intervall, innerhalb dessen die Messungen liegen, und zum Zweck korrekter Wiedergabe ist dies Intervall zu ermitteln, und unter Berücksichtigung der Art der Gewinnung der Zahlen darzustellen.

Viel einfacher gestaltet sich daher die Zeichnung, wenn die Tabelle selbst die Intervalle angibt, zwischen denen die Messungen liegen. Es sind dann diese Grenzwerte, vom Maximum angefangen, als Ordinaten, und die Summe der jeden Grenzwert überschreitenden Objekte als Abszissen zu nehmen, und die so gefundenen Punkte zu verbinden.

Ist zugleich der Durchschnittswert der Messungen jeder Gruppe angegeben, oder läßt er sich aus der Summe der Messungen oder sonst ermitteln, so muß die Figur ihm entsprechen. Es sind dann zwei Figuren zu einer einzigen zu vereinigen, die Treppenfigur der Durchschnitte und die polygonförmige Figur der Intervallgruppen. Von diesen sind zuverlässig gegeben die Grenzpunkte der Intervallgruppe und die Flächeninhalte ihres Durchschnitts. Die Höhen der hinzuzufügenden und hinwegzunehmenden Dreiecke stehen fest; teilt man die oberen Seiten der Rechtecke im umgekehrten Verhältnis der angrenzenden Höhendifferenzen und verbindet die Teilpunkte mit den Mittelpunkten, so werden zwischen je zwei Ordinaten gleiche Dreiecke hinzugefügt und hinweggenommen. Die Figur entspricht sowohl der Reihe der Durchschnitte als der Reihe der Intervallgruppen.

In allen Fällen stimmt die Kumulativ- mit der Merkmalskurve genau überein; die von beiden eingeschlossenen Flächen ergänzen sich zu einem Rechteck.

II. Häufigkeitskurve

In ähnlicher Weise ist die Auflösung der Messungsgruppen in Intervallgruppen für die Häufigkeitskurve von Bedeutung, diese ist sogar nur unter dieser Voraussetzung darstellbar. Man kann zwar für jede bestimmte Größe der Messung die dazugehörige Zahl der Objekte durch eine Ordinate darstellen; für die dazwischen liegenden Werte der Abszisse wäre aber diese Ordinate gleich 0, und man würde nur die Zahl der Objekte darstellen, nicht die Häufigkeit im Sinn der Häufigkeitskurve, die die Beziehungszahl zwischen dieser Zahl und der Ausdehnung des Intervalls ist. Wollte man diese Beziehungszahl für genaue Messungen durch eine Kurve darstellen, so würden alle Objekte auf dasjenige unendlich kleine Intervall entfallen, das die Messung selbst enthält; die Häufigkeit wäre für dieses unendlich groß, und dazwischen gleich 0. Nur wenn man sich die möglichen Messungen über ein Intervall verteilt denkt,

[1]) Man kann auch die angrenzenden Stücke halbieren, ferner kann man die ganzen oder halben Stücke vertauschen, dann bezeichnet der Schnittpunkt der Verbindungslinie unmittelbar die gewünschte Teilung im umgekehrten Verhältnis.

kann man also für dieses die Häufigkeit durch eine Kurve darstellen. Die Zahl der Objekte ist dann durch den eingeschlossenen Flächeninhalt, das Intervall durch die Breite und daher die Häufigkeit durch die durchschnittliche Höhe dargestellt.

Hat man demgemäß die kumulative Figur aus schrägen Linien zusammengesetzt, so entspricht ihr eine einfache Häufigkeitskurve, deren Ordinate jeweils den Wert hat, um den die kumulative Kurve bei Beibehaltung der Steilheit in der Einheit der Intervalle zunehmen würde.[1]) Sie verläuft also für jedes gradlinge Stück der kumulativen Kurve horizontal, geht dann bei Änderung der Steilheit plötzlich zu einer größeren oder kleineren Ordinate über, um wieder horizontal zu verlaufen. Der Polygonform der kumulativen Kurve entspricht also eine Treppenform für die einfache Häufigkeit; letztere wird durch eine Reihe von Rechtecken ausgedrückt. Diese beginnen bei Messungsgruppen mit gleichen Intervallen ein halbes Intervall vor der Abszisse für jede Messung, und gehen ein halbes Intervall über diese hinaus. Der Flächeninhalt der ganzen Figur beträgt soviel mal ein Rechteck mit der Abszisse und der Ordinate für 1 als Seiten, als Objekte gezählt worden sind. Die Höhe jedes der Rechtecke ist proportional der Strecke, um die die Ordinate der kumulativen Kurve innerhalb seiner Breite zugenommen hat. Dies gilt auch dann, wenn man die Rechtecke in beliebig viele senkrechte Streifen teilt, sowie wenn man sie von dem ersten angefangen summiert; dann ist ihre jeweilige Summe der entsprechenden Ordinate der kumulativen Kurve proportional.

Daraus folgt, daß man auch umgekehrt die kumulative Kurve aus der Häufigkeitskurve konstruieren kann, und zwar auch dann, wenn diese nicht aus Rechtecken besteht. Denn der Flächeninhalt einer jeden Figur ist praktisch gleich der Summe unendlich schmaler Trapeze, in den man ihn zerlegt denken kann, und diese sind gleich Rechtecken mit gleicher Breite und der Höhe gleich der durchschnittlichen Höhe der Trapeze.

Der wichtigste Fall ist der, daß die Häufigkeitskurve aus graden, schrägen Linien besteht, also ein Polygon bildet. Eine solche Form nimmt die Häufigkeitskurve an, wenn sie der Ausgleichung unterworfen wird.[1])

Graphische Ausgleichung

Während die genaue Darstellung der Ergebnisse das erste Ziel eines Diagramms und der Ausgangspunkt für alle weiteren Berechnungen und Konstruktionen sein muß, ist es für manche Zwecke doch notwendig, die wesentlichen Züge einer Zahlenreihe von dem unwesentlichen zu trennen. Nur aus ihnen läßt sich das zugrunde liegende Gesetz klar erkennen, lassen sich Vergleiche mit anderen Reihen wirklich beobachteter Zahlen und mit solchen Zahlenreihen anstellen, die aufgrund einer vermuteten stetigen Beziehung zweier Größen gebildet sind.

Der Ausgleich soll die Kontinuität herstellen, er soll die kleinen Schwankungen beseitigen, ohne die Grundtendenz der Kurve und die Hauptergebnisse der Reihen zu verändern.

1. bei gleichen Intervallen

Um dies zu erreichen, kann man bei gleichen Intervallen und dementsprechend gleicher Breite der Rechtecke der Häufigkeitskurve die Mittelpunkte

[1]) Die Häufigkeitskurve entspricht dem Differenzialquotienten der kumulativen Kurve, d. h. der durch jedes einzelne Stück derselben dargestellten Funktion.

[1]) Im Folgenden sollen nur elementare Ausgleichsmethoden besprochen werden, die sich zur unmittelbaren Konstruktion eignen.

ihrer oberen Seiten verbinden. Dann werden gleiche Dreiecke hinzugefügt und weggenommen; der Inhalt der von der Gesamtkurve eingeschlossenen Fläche bleibt also unverändert; dagegen werden die einzelnen Häufigkeiten = 0, die man zum Zweck der Vollständigkeit des Ausgleichs an beiden Enden der Kurve in gleichem Abstand hinzuzufügen hat, oder die sich etwa innerhalb der Reihe der Rechtecke befinden. Die Breite der Figur wird dadurch auf beiden Seiten um ein halbes Intervall vergrößert. Da jedoch die Dreiecke von jedem der Rechtecke symmetrisch nach links und rechts verschoben worden sind, ist der durchschnittliche Grad des Merkmals unverändert geblieben und im einzelnen nur innerhalb enger Grenzen verschoben worden. Die Ausgleichung ist daher als korrekt anzusehen.

Eine Ausnahme ist nur dann zu machen, wenn es tatsächlich unmöglich ist, daß ein Objekt kleiner als die kleinste, oder größer als die größte angegebene Messung ist. Die Abweichungen können dann nur positiv bezw. negativ sein; die Reihe von unmeßbar verschiedenen Graden des Merkmals, als deren Repräsentant die Messung erscheint, hat nur die halbe Ausdehnung, ihre Häufigkeit ist daher streng genommen durch ein Rechteck von halber Breite und doppelter Höhe auszudrücken.

Weitere Ausgleichung

Will man die Ausgleichung verstärken, so kann man die Verbindung der Mittelpunkte wiederholen, und zwar liegen die Mittelpunkte der ersten schrägen Strecke auf den Mittelpunkten der senkrechten Strecken der Treppenfigur, und alle Mittelpunkte abwechselnd auf den Ordinaten für die Messungen und denjenigen für das halbe Intervall; bezw. auf den Grenzen und in der Mitte eines Intervalls der Intervallgruppen.

Bei gleichen Intervallen kann man also die Häufigkeitskurve in beliebiger Ausgleichung leicht aus den Rechtecken herstellen. Diese brauchen dann auch nicht aus der kumulativen Kurve hergeleitet zu werden, sondern ergeben sich unmittelbar aus den Daten. Sind die Zahlen der Objekte für Messungen mit gleichen Intervallen angegeben, so sind diese, als Ordinaten auf der Abszisse des betreffenden Merkmals errichtet, zugleich die Mittellinien der Rechtecke der Häufigkeit in bezug auf diese Größe der Intervalle. Die Verbindungslinie ihrer Endpunkte und der in gleichen Intervallen benachbarten Punkte für die Häufigkeit 0 bildet die erste ausgeglichene Kurve.

Sind die Zahlen der innerhalb gleicher Intervalle liegenden Objekte gegeben, so liegen die Rechtecke zwischen diesen Ordinaten, und für die erste Ausgleichung sind die Mittelpunkte ihrer oberen Seite zu verbinden.

Die zweite Ausgleichung ist direkt zu erhalten, indem man die Mittelpunkte der senkrechten Strecken der Figur, d. h. die Durchschnitte der Häufigkeiten zweier benachbarten Intervalle auf der Ordinate ihrer Grenze aufträgt.

2. bei ungleichen Intervallen

Sind die Intervalle ungleich, so ist die Häufigkeit zu berechnen, die sich ergeben würde, wenn sie gleiche Größe hätten. Man muß dann eine beliebige Größe als Normalintervall annehmen, und die Zahl der Objekte im Verhältnis vermehren, in dem das Normalintervall zu dem betreffenden Intervall steht. Setzt man dieses Normal-Intervall = 1, so hat man die Zahl der Objekte durch die Größe des Intervalls zu dividieren; z. B.

	10	Objekte	1—2
	20	„	2—4
	10	„	4—6
ergibt	10	„	1—2
	10	„	2—3
	10	„	3—4
	5	„	4—5
	5	„	5—6

oder bei Messungsgruppen

10 Objekte zu 1		10 Objekte 0—2
20 „ „ 2		20 „ 1—3
10 „ „ 4		10 „ 3—5
das ist	5 Objekte	1—0
	15 „	1—2
	10 „	2—3
	5 „	3—4
	5 „	4—5

Bei ungleichen Messungsintervallen schieben sich also die Rechtecke der Häufigkeitskurve in ähnlicher Weise übereinander wie die Dreiecke der kumulativen Kurve, und bei gleichen Verschiebungen entsprechen diese Kurven einander genau. Dabei ist die Häufigkeitskurve leichter zu zeichnen, da Rechtecke leichter zu addieren sind als Dreiecke. Bevor man diese Rechtecke weiter ausgleicht, wird man sie zweckmäßig auf gleiche Intervalle reduzieren, indem man die Abszissenachse in gleiche Teile teilt, und die dazwischenliegenden Rechtecke in ein einziges verwandelt.

Man kann auch die erste Ausgleichung direkt zeichnen, indem man in den Abszissen der gemessenen Werte Ordinaten entsprechend der Menge der Objekte errichtet, und um diese gleichseitig Dreiecke errichtet, die diese Ordinaten zur Höhe und die doppelte Breite des Normal-Intervalls zur Grundlinie haben. Ihr Inhalt ist dann gleich der Zahl der Objekte multipliziert mit dem Intervall 1. Sie stellen somit die Ausgleichung richtig dar. Addiert man alle Ordinate dieser Dreiecke, so erhält man eine fortlaufende Kurve. Will man diese ausgleichen, so kann man die Höhe der Dreiecke halbieren, die Grundlinie verdoppeln usw.

Sind Intervallgruppen mit ungleichen Intervallen gegeben, verhält sich die Häufigkeit (h) innerhalb jedes Intervalls, zur Zahl der Objekte (z), wie das Intervall 1 (1) zur Breite des Intervalls (b). Man hat demgemäß

da $h : z = 1 : b$

und $h = z : b$

die Zahl der Objekte durch die Breite des Intervalls zu dividieren, um die Häufigkeit zu erhalten.

Zum Ausgleich kann man die entstandenen Rechtecke in der angegebenen Weise auf Intervalle reduzieren, oder sie durch Dreiecke von gleicher Höhe und doppelter Breite ersetzen und deren Ordinaten addieren.

[1]) Man kann zwei Rechtecke in eines verwandeln, das die Summe der beiden andern zur Grundlinie hat, d. h. zwei Treppenstufen in eine, indem man die Differenz der Höhen im umgekehrten Verhältnisse der Grundlinien teilt. Zu diesem Zweck verlängert man in Gedanken die wagrechten Strecken bis zu den Schnittpunkten mit den äußeren Ordinaten der beiden Treppenstufen und verbindet diese Schnittpunkte mit einander. Diese teilt die mittlere Ordinate im umgekehrten Verhältnis der anliegenden wagrechten Strecken.

Man kann auch die Mittelpunkte der wagrechten Strecken zunächst durch gedachte Linien verbinden und deren Schnittpunkte mit den senkrechten Linien notieren; diese werden dann im Verhältnis der anstoßenden Strecken geteilt. Vertauscht man diese Teile und verbindet die Grenzpunkte mit den Mittelpunkten der wagrechten Strecken, so erhält man Dreiecke, deren Grundlinien sich umgekehrt verhalten wie die Höhen, deren Flächen also gleich sind. Von diesen kann das eine zugefügt, das andere hinweggenommen werden, ohne die Fläche der Kurve zu verändern.

Der polygonförmigen Häufigkeitskurve entspricht eine aus Parabelstücken zusammengesetzte kumulative Kurve

Hat man auf eine der angegebenen Weisen die Häufigkeitskurve aus schrägen Linien zusammengesetzt, und will die ihr genau entsprechende kumulative Kurve konstruieren, so wird diese aus Parabelstücken zusammengesetzt sein. Das erste Stück der Häufigkeitskurve schließt ein Dreieck ein; teilt man das Intervall in z. B. 10 Teile, so verhalten sich die eingeschlossenen Flächen bis zu den Abszissen a, b, c, wie ½ a^2 zu ½ b^2 zu ½ c^2 und bis 0, wie 0

0,1 wie ½ . $0{,}1^2$ = ½ . 0,01 = 0,005
0,2 „ ½ . $0{,}2^2$ = ½ . 0,04 = 0,020
0,3 „ ½ . $0{,}3^2$ = ½ . 0,09 = 0,045
: „ ½ . $0{,}4^4$ = ½ . 0,16 = 0,080
0,9 „ ½ . $0{,}9^2$ = ½ . 0,81 = 0,405
1,0 „ ½ . $1{,}0^2$ = ½ . 1 = 0,500

Die Ordinaten der kumulativen Kurve verhalten sich also wie die Quadrate der Abszissen, d. h. die Punkte liegen auf einer Parabel. Die obigen Werte sind mit der Höhe des Dreiecks, d. h. mit der erreichten Häufigkeit, zu multiplizieren; bei 1,0 erreicht die kumulative Kurve dieselbe Höhe, wie wenn anstelle des Dreiecks ein Rechteck mit gleicher Fläche die Häufigkeit ausgedrückt hätte.

Dem folgenden schrägen Stück der Häufigkeitskurve entspricht der Form nach ein Stück einer Parabel, deren Achse bei dem Schnittpunkt der Verlängerung dieses Stückes mit der Abszissenachse begonnen und deren Ordinaten dem jeweiligen Flächeninhalt des eingeschlossenen Dreiecks proportional sind. Ist das schräge Stück steiler, so wäre das von ihr eingeschlossene Dreieck kleiner und das Parabelstück ist aufwärts so weit zu verschieben, daß es sich an das vorhergehende anschließt. Hat ein schräges Stück die Richtung nach abwärts, so steigt die kumulative Kurve in abnehmendem Maße an. Es ist dann ein Stück einer Parabel zu zeichnen, die über dem Schnittpunkt des schrägen Stückes mit der Abszissenachse ihren Höhepunkt erreichen würde. Für praktische Zwecke würde es genügen, davon auszugehen, daß die kumulative Kurve am Ende jedes schrägen Stückes denselben Wert erreicht, den sie bei einem Rechteck von gleicher Fläche wie das eingeschlossene Trapeze auch erreicht hätte, und ebenso nach der Hälfte der schrägen Strecke. Ersteres Rechteck hat die Höhe der mittleren Ordinate der schrägen Strecke, letzteres diejenigen der mittleren Ordinate ihrer ersten Hälfte, beide haben die halbe Breite des Intervalls; die kumulative Kurve steigt also von dem gegebenen Anfangspunkt bis zur Mitte des Intervalls, um ¼ der Differenz der Ordinaten.

Nach Bedarf kann man das schräge Stück auch z. B. in vier gleiche Teile teilen, dann entstehen vier Trapeze mit gleichmäßig zu- bezw. abnehmendem Inhalt. Die kumulative Kurve wächst um Strecken, die stets den Flächeninhalt des betreffenden Trapezes proportional sind. So erhält man drei Punkte, durch die man aus freier Hand ein Parabelstück annähernd genau zeichnen

kann. Hat z. B. die Häufigkeitskurve am Anfang eines Intervalls die Ordinate 4, am Ende 8, so nimmt die kumulative Kurve bis zur Hälfte um ½ . $\frac{4 + \frac{4+8}{2}}{2}$ = ½ . 5 zu, von da bis zum Ende um ½ . $\frac{\frac{4+8}{2} + 8}{2}$ = ½ . 7. zusammen um ½ . 5 + ½ . 7 = 6 = $\frac{4+8}{2}$ zu.

Anordnung der Achsen

Es bleibt noch zu erwähnen, daß die Anordnung der Figur nach Bedarf verändert werden kann. Man kann z. B. die Merkmalskurve mit der niedrigsten Messung beginnen lassen, so daß sie dieselbe ansteigende Richtung erhält, wie eine Kurve zeitlicher zunehmender Entwicklung. Die dazugehörige kumulative Kurve bewegt sich dann in absteigender Linie. Ferner kann man, von dem sonstigen Gebrauch abweichend, die abhängige veränderliche Größe durch die Abszissen ausdrücken, und so diejenige Figur, die ich bisher kumulative Figur genannt habe, Merkmalskurve zu nennen, und umgekehrt.

Endlich sind oft Tabellen so angelegt, daß sie zwar Intervallgruppen enthalten, deren eine Grenze aber zugleich eine der runden und daher vermutlich stark vertretenen Zahlen für die Messungen ist. Dann liegt der Schwerpunkt jeder Gruppe nicht in ihrer Mitte. Falls sich nicht die Durchschnittseigenschaft jeder Gruppe oder der ganzen Reihe ermitteln läßt, ist die Grenze der Intervalle schätzungsweise ein wenig zu verschieben; z. B. Gruppe 10—10,9 als 9,9—10,9 anzusetzen.

Zusammenfassung

Faßt man das bisherige zusammen, so ergibt sich, daß die Kurve der Messungen des Merkmals und die kumulative Kurve bei genauer Zeichnung stets übereinstimmend und die Ordinaten der kumulativen Kurve den von der Häufigkeitskurve eingeschlossenen Flächen entsprechen und zwar stimmt die Polygonform der kumulativen Kurve mit der Treppenform der Häufigkeitskurve überein und eine parabolische kumulative Kurve mit einem Polygon der Häufigkeitskurve.

Die Merkmals- und kumulative Kurve ist aus jeder Tabelle unmittelbar zu entnehmen, und zwar bei Messungsgruppen in Treppenform, bei Intervallgruppen in Polygonform. Die Treppenform läßt sich leicht in ein Polygon verwandeln, wenn die Intervalle oder die Mengen gleich sind; dann sind die Mittelpunkte der gleichen Strecken zu verbinden. Sind sowohl Intervalle als Mengen ungleich, so sind die Intervalle zu halbieren und die Mengen im umgekehrten Verhältnis der anstoßenden Strecken zu teilen, eventuell ebenfalls zu halbieren.

Die Häufigkeitskurve ist nur dann leicht zu zeichnen, wenn die Intervalle gleich sind; bei Messungsgruppen ist in jeder Messung eine Ordinate zu errichten, die die Mittellinie der Treppenstufe von gleicher Höhe bildet. Verbindet man diese unter sich und mit den benachbarten Intervallen mit der Häufigkeit = 0, so erhält man als erste Ausgleichung ein Polygon, durch weitere Verbindung der Mittelpunkte erhält man die zweite, dritte und folgende Ausgleichung. Bei Intervallmengen mit gleichen Intervallen ist die Ordinate in der Mitte jedes Intervalls zu errichten (das Rechteck füllt das Intervall aus), im übrigen ist ebenso zu verfahren.

Bei ungleichen Intervallen ist die Herstellung schwieriger. Bei Messungsgruppen kann man für jede Messung die Zahl der Objekte als Ordinate und

Höhe eines gleichseitigen Dreiecks von der doppelten Breite des als normal angenommenen Intervalls nehmen, und die Ordinaten dieser Dreiecke addieren. Dann erhält man die erste Ausgleichung, die aber nur durch Verbreiterung der Dreiecke in die folgenden verwandelt werden kann.

Bei Intervallgruppen kann man die Zahl der Objekte durch das Intervall dividieren und so die Höhe von Rechtecken bestimmen, deren Breite gleich dem Intervall, deren Fläche also gleich der Zahl der Objekte ist. Die entstehende Treppenfigur entspricht genau der polygonförmigen kumulativen Kurve der Intervallgruppen; sie ist daher aus dieser herzustellen, und zwar, indem man für jedes der ungleichen Intervalle als Ordinate den Zuwachs nimmt, den die Häufigkeitskurve erreicht hätte, wenn sie diese Richtung ein Normal-Intervall hindurch beibehalten hätte. Diese Ableitung läßt sich auch auf die zur Polygonform ausgeglichene kumulative Kurve der Messungshäufigkeiten anwenden. Zur weiteren Ausgleichung empfiehlt es sich, die Rechtecke innerhalb jedes Normalintervalls in ein einziges zu verwandeln.

Einfacher, wenn auch weniger genau ist es, die von der kumulativen Kurve bei jedem Normal-Intervall erreichten Punkte zu verbinden, und sie so zu einem gleichmäßigen Polygon zu gestalten, aus dem die Häufigkeitskurve leicht abgeleitet werden kann, und auch die Ausgleichungen einfach bewirkt werden können.

Aus dieser Übersicht folgt, daß die Häufigkeitskurve nur bei gleichen Intervallen leicht herzustellen ist, bei ungleichen dagegen leichter aus der kumulativen Figur hergeleitet werden kann; ebenso ist die Ausgleichung dann erleichtert.

Anwendung auf andere einander entsprechende Kurven

Die hier aufgestellten Regeln genauer Konstruktion und Ausgleichung haben ihre Bedeutung nicht nur für die Übereinstimmung der kumulativen und Häufigkeitskurve, sondern allgemein für alle Fälle, in denen die Ordinate der einen Kurve die Summe des Flächeninhaltes der anderen Kurve bis zur gleichen Abszisse darstellt. Insbesondere können sie auf den Fall entsprechend angewendet werden, daß Bestands- und Bewegungsmassen in ihrer gegenseitigen Abhängigkeit dargestellt werden sollen, z. B. die Zunahme der Bevölkerung und ihr jeweiliger Bestand. Die Treppenfigur der Zunahme entspricht die Polygonform des Bestandes und der Polygonform der absoluten Zunahme pro Jahr oder Jahrzehnt, die aus Parabelstücken zusammengesetzte Kurve ihres Betrages. Denn einer gleichbleibenden Zunahme entspricht gleichmäßiges Ansteigen, einer zunehmenden Zunahme ebensolches Ansteigen. Der Intensität der Zunahme entspricht die Ordinate, ihrem Betrag in einer Zeitstrecke die Fläche.

Vorzüge der kumulativen bezw. Häufigkeitskurve

Bei der Verwendung der Kurven ergeben sich ebenso wie bei der Herstellung Vorzüge für die eine oder andere Art und zwar ist hier der erstrebte Zweck maßgebend.

Will man die Grade der Eigenschaften und ihre Verteilung bei Reihen verschiedener Gegenstände oder gleichartiger Gegenstände zu verschiedenen Zeiten

[1] Eine andere Anwendung ist die auf das Verhältnis zwischen Beschleunigung und Schnelligkeit, sowie zwischen Schnelligkeit und zurückgelegter Strecke. Bei gleichbleibender Beschleunigung nimmt die Geschwindigkeit gleichmäßig zu, der zurückgelegte Weg im Verhältnis der Quadrate der Zeit, die dazugehörige Kurve ist eine Parabel, wie aus der Flugbahn im luftleeren Raum zu ersehen ist. Vergl. Anhang zu A.

vergleichen, so ist zu unterscheiden, ob man auch die absoluten Zahlen oder nur ihre Gruppierung vergleichen will. Im letzteren Fall kann man die Merkmals- bezw. kumulativen Kurven so konstruieren, daß die Gesamtzahlen der Objekte durch größte Abszissen bezw. Ordinaten von gleicher Länge dargestellt werden; z. B. indem man die Gesamtzahl = 100 setzt und die auf jede Gruppe entfallenden Objekte in Prozent der Gesamtzahl ausdrückt. Wenn man diese Berechnung auf die Häufigkeitskurve anwendet, wird dagegen erreicht, daß die von diesen eingeschlossenen Flächen gleich groß werden, insofern sind beide Arten von Kurven gleich gut zur Vergleichung geeignet.

Wenn man aber die so auf gleiche Gesamtzahlen reduzierten Kurven über- oder nebeneiander legt, so werden die unausgeglichenen Häufigkeits-Kurven noch keinen Vergleich erlauben, weil sie in der Regel stark oszillieren, und zwar um so mehr, je kleiner die Intervalle sind, und je weniger durch eine große Zahl von Beobachtungen zufällige Abweichungen ausgeglichen sind. Die einfache Häufigkeitskurve verlangt also entweder ein sehr großes Material, was dadurch um so weniger homogen sein kann, oder die Bildung größerer Gruppen, oder endlich eine nachträgliche Ausgleichung. Dadurch werden gerade die Feinheiten verwischt. Außerdem bringt die Gruppenbildung und Ausgleichung ein Element der Willkür mit sich. Eine Kurve kann eine ganz verschiedene Gestalt annehmen, je nachdem das Urmaterial gruppiert ist, und je nachdem infolgedessen die Anhäufungsstellen auf dieses oder jenes Intervall entfallen, und gleichmäßig oder ungleichmäßig verteilt sind. Ebenso bringt die Ausgleichung sehr verschiedene Wirkungen hervor, je nach der Methode und dem Grade, in dem sie durchgeführt ist.

Die kumulative Kurve dagegen bedarf keiner Ausgleichung, um vergleichbar zu sein. Sie kann nicht auf- und abschwanken, sondern sich nur mit wechselnder Steilheit in derselben Richtung bewegen. Sie hat bereits eine Ursache der Ausgleichung in sich, indem die Vermehrungen und Verminderungen der einzelnen Gruppen sich auf die folgenden übertragen und dadurch gegenseitig abschwächen. Infolgedessen wird sie von denselben Methoden der Ausgleichung, die Häufigkeitskurve total umgestalten, nur viel schwächer beeinflußt, so daß es schwer ist, die stärkeren und schwächeren Ausgleichungen auf derselben Figur unterscheidbar zu zeichnen.

Der Vorteil der kumulativen Kurve liegt also darin, daß sie einer Ausgleichung zum Zweck der Vergleichung nicht bedarf; will man eine Ausgleichung vornehmen, so ist die Häufigkeitskurve vorzuziehen.[1]

2. Die kumulative Kurve läßt den Zentralwert, perzentile und andere Grade erkennen

Ein weiterer Vorzug der kumulativen Kurve besteht darin, daß man aus ihr die quartilen, dezilen und perzentilen Grade und den Zentralwert konstruktiv bestimmen kann (vgl. Zizek, S. 258), ohne das Urmaterial zu benutzen. Man teilt zu diesem Zweck die größte Ordinate der kumulativen Kurve (bezw. die größte Abszisse der Eigenschaftskurve) in 2, 4, 10, 100 Teile, errichtet Senkrechte und bezeichnet die Schnittpunkte mit der Kurve.

Wollte man aus einer Tabelle, die nur die Zahl der Objekte nach Eigenschaften angibt, den Zentralwert und die Quadrilen usw. berechnen, so müßte man ebenfalls eine kumulative Reihe bilden und unter der Annahme, daß die Objekte in jedem Intervall gleichmäßig verteilt sind, den vermutlichen Wert

[1] Siehe Anhang.

der Quartilen usw. bestimmen; das Ergebnis wäre dasselbe, wie bei der Konstruktion. Letztere ist aber leichter und anschaulicher.

Umgekehrt kann man diese Werte durch bloßes Abzählen aus dem Urmaterial entnehmen, indem man einfach diejenigen Objekte oder Paare von Objekten herausgreift, die auf der Grenze der Hälften, Viertel, Zehntel oder Hundertstel der ganzen Reihe stehen.

Die Bedeutung dieser Werte besteht darin, daß sie ein vereinfachter und leicht vergleichbarer Ausdruck der ganzen Reihe sind. Sie haben untereinander gleiches Gewicht, können also den Berechnungen zugrunde gelegt werden.[1])

Aus den Kurven kann man ferner eine Reihe von Beziehungen der Glieder der Reihe teils unmittelbar ersehen, teils auf sie schließen, wenn man sich die negativen oder die reziproken Werte der dargestellten Größen durch Schätzung, Rechnung oder Konstruktion bildet.

Je steiler die kumulative Kurve, je flacher die Merkmalskurve verläuft, desto größer ist die Dichtigkeit der aufeinanderfolgenden Werte oder die Häufigkeit. Der Zahl der Objekte pro Intervall entspricht für jede Ordinate der kumulativen Kurve dem Wert um den die Ordinate zunehmen würde, wenn die Kurve ihre Richtung bis zu der Ordinate beibehalten würde.

Die Häufigkeitskurve stellt diese Dichtigkeiten konstruktiv dar, und läßt so die Verteilung der Häufigkeit über die Intervalle der Merkmale deutlicher erkennen, als es die Schätzung von Winkeln ermöglichen würde. Man sieht unmittelbar, wo der dichteste Wert und die anderen Anhäufungsstellen liegen, und kann die Häufigkeit leicht vergleichen.

Die Werte, um die die Ordinaten der Merkmarlskurve bei gleichen Abständen zu- bezw. abnehmen, entsprechen den Differenzen der Merkmale benachbarter Glieder. Man sieht sofort, ob diese Differenzen im Beginn der Reihe bei den höchsten Werten der Eigenschaft, in der Mitte oder an einer anderen Stelle der Reihe am größten sind. Will man sich diese Differenzen verdeutlichen, so kann man in derselben Weise eine Kurve aus der Merkmalskurve bilden, wie die Häufigkeitskurve aus der kumulativen Kurve entsteht. Man berechnet dann, wie groß die Differenz der Merkmale würde, wenn die Merkmalskurve ihre Richtung für einen bestimmten Teil, z. B. $\frac{1}{4}$, $\frac{1}{10}$, $\frac{1}{00}$ oder die ganze Reihe beibehielte. Diese Unterschiede sind wieder trigonometrisch als Tangenten des Winkels der Merkmalskurve multipliziert mit einer Konstante, auszudrücken, die hier eine bestimmte Zahl von Objekten bedeutet.

Vergleich der Häufigkeitskurve einer empirischen Reihe mit der Fehlerkurve

Die Häufigkeitskurve dient nun nicht nur zur Veranschaulichung der Messungsresultate, sondern auch zur Prüfung derselben, daraufhin, ob sie dem Gauß'schen Fehlergesetz entsprechen; ob also die Abweichungen von dem häufigsten Wert so gruppiert sind, wie es nach der Wahrscheinlichkeitsrechnung die Resultate unendlich vieler und kleiner, zufällig verteilter Veränderungen sein würden. (Siehe Fig. 7.)

Entspricht die Kurve dem Gauß'schen Fehlergesetz, so kann man wohl das Maß der Abweichung bestimmen, erhält jedoch keinen Aufschluß über den

[1]) Der Zentralwert ist zugleich derjenige, von welchem aus die Summen aller positiv genommenen Abweichung die kleinste ist, während von arithmetischen Mitteln aus die Summe der Quadrate der Abweichungen ein Minimum ist.

Anmerkung: Ist der Winkel der Kurve mit der Abzissenachse = x; das Intervall, auf das sich die Häufigkeit bezieht, (Normalintervalle) = i, so ist die Häufigkeit = i tang x.

Charakter der wirkenden Ursachen. Die Reihe zeigt deutlich den Typus der Erscheinung; von den Abweichungen weiß man nur, daß sie nicht aus solchen Ursachen hervorgehen, deren Wirkung eine erkennbare Veränderung derjenigen Gruppierung hervorbringt, die bei unendlich vielen zufälligen Ursachen eintreten würden.

Weiht die Häufigkeitskurve von der Fehlerkurve ab, so kann dies verschiedene Ursachen haben.

Ursachen der Verschiedenheit

Es kann sein, daß nicht eine einheitliche Reihe, sondern zwei oder mehr vorliegen, deren Summe in der Häufigkeitskurve erscheint. Liegen ihre Gipfel genügend weit auseinander, so daß an dem Schnittpunkt die Biegung nach aufwärts überwiegt, so hat die aus der Addition der entsprechenden Glieder und von deren Ordinaten entstehende Kurve zwei Gipfel; liegen die Gipfel der Teilkurven näher beieinander, so ist wenigstens der Gipfel der Gesamtkurve breiter, als der einer einfachen Fehlerkurve. Ein Beispiel der Anwendung dieses Gedankens ist die Auflösung der Häufigkeitskurve der Rekrutenmaße in Italien in zwei Type, die auf Rassenverschiedenheit zurückgeführt werden. [1])

Eine ähnliche Ursache haben Abweichungen von dem normalen Verlauf der Häufigkeitskurve, die auf eine in verschiedenem Grade erfolgte Mischung und Beeinflussung zweier Typen zurückgeführt werden können. Dabei kann man sowohl an biologische Rassenmischung denken, als an eine bloß soziale oder sozusagen statistische Mischung, wenn nämlich eine Erscheinung nur in Durchschnitten oder Beziehungszahlen zum Ausdruck kommt. Es läßt sich zeigen, daß die Wahrscheinlichkeit durch eine zufällige Mischung und Gruppierung verschiedener Elemente eine dem Gesamtdurchschnitt entsprechende Zusammensetzung zu erhalten, um so größer wird, je größer die Gruppen sind. Ebenso ist die Wahrscheinlichkeit, daß sich Beziehungszahlen für Teilgruppen der Beziehungszahl für das Ganze nähern, um so größer, je größer die Teilgruppe im Verhältnis zum Ganzen ist. Ordnet man daher Beziehungszahlen für Gruppen von verschiedenem Umfang, so wird die nach dem Gauß'schen Gesetz zu erwartende Kurve nur dann wirklich dem Gesetz des Zufalls entsprechen, wenn die Bildung der Gruppen rein zufällig erfolgt. Das Gauß'sche Gesetz ist deshalb zur Beurteilung von Reihen von Durchschnitten und Beziehungszahlen nicht unmittelbar anwendbar; vielmehr ist es notwendig, die durch die Gruppierung hervorgerufenen Einflüsse zuerst zu beseitigen. Dies gilt jedoch nur insoweit, als der sachliche Inhalt dieser Durchschnitte und Gruppierungszahlen beurteilt werden soll. Um über die Art der Gruppierung selbst einen Aufschluß zu gewinnen, ist es sehr wohl zulässig, sie mit einer solchen zu vergleichen, die bei rein zufälliger Mischung und Zusammenfassung entstehen würde, und dann kann unter Umständen das Gauß'sche Gesetz einen Vergleichsmaßstab abgeben.

Ungleichmäßiger Erfolg der abändernden Ursachen

Eine dritte Art der Abweichung von der Fehlerkurve kann bei der Häufigkeitskurve dadurch entstehen, daß die abändernden Ursachen nicht gleichmäßig wirken, sondern in verschiedenem Maße, je nachdem eine Abweichung bereits erfolgt ist. Der einfachste Fall ist derjenige, daß das Maß der weiteren Abänderung dem Maß proportional ist, den der betreffende Gegenstand erreicht hat. So ist das absolute Wachstum einer Stadt, einer Bevölkerung vor allem

[1]) Benini, Principi di demografia.

durch die Größe bestimmt, die diese Bevölkerung erreicht hat. Ein Zuwachs von 10 000 Personen für eine Bevölkerung von 1 Million bedeutet dasselbe, wie 1000 Personen Zunahme für 100 000, nämlich einen Zuwachs von 1%. Diese Art von Ursachen bewirken also nur dem absoluten Betrag nach verschiedene, relativ aber gleiche Abänderungen. Um diese relative Gleichheit graphisch zur Erscheinung treten zu lassen, kann man sich der Logarithmen der Maßzahlen bedienen, und die so erhaltene Kurve mit der Fehlerkurve vergleichen.

Andere Ursachen wirken um so schwächer, je größer der bisher erreichte Erfolg ist, sei es, weil sie die Gegenkräfte erzeugen, die ihre weitere Wirkung hemmen, sei es, weil sie nur einmal in jedem Fall wirken können und jeder Erfolg die Zahl der Möglichkeiten weiterer Erfolge verringert. Diese Erscheinung kann mit einer Grundtendenz geometrischer Progression verbunden sein. So ist die Wahrscheinlichkeit für die Verbreitung einer ansteckenden Krankheit ceteris paribus sowohl von der Zahl der bereits Befallenen, als der noch Ansteckungsfähigen abhängig. In jedem Kreise von Personen, die einander der Gefahr der Ansteckung aussetzen können, ist die Wahrscheinlichkeit (W), daß eine Person aus der Zahl (A) der Ansteckungsfähigen mit einer solchen aus der Zahl (B) der von der Krankheit Befallenen zusammentrifft.

$$W = \frac{A \cdot B}{(A+B)\,(A+B-1)}$$

Solange B noch sehr klein ist, wird A kaum merklich vermindert, und die Progression ist fast geometrisch; die relative Zunahme von B nimmt jedoch in steigendem Maße ab, während die absolute solange zunimmt, bis A = B geworden ist, um dann symmetrisch ebenfalls bis auf 0 zu sinken.

Es lassen sich die mannigfaltigsten Formen denken, in denen die Größe der Abänderung von dem bisher erreichten Betrag abhängig ist; man kann darnach Regeln eines Glückspiels bestimmen, nach denen eine solche Wahrscheinlichkeit a priori gegeben ist. Für empirische Reihen kann man aber nicht von vorneherein angeben, welches in dem einzelnen Fall die Beziehung zwischen der Änderung und dem bisher erreichten Maß sein wird. Setzt man aber voraus, daß die Zahl der Änderungen nach dem Gesetz des Zufalls bestimmt wird, so läßt sich a posteri ermitteln, wie groß der Betrag der Abweichungen gewesen sein muß. Man ordnet die Objekte nach den Graden, in denen sie das Merkmal besitzen und teilt sie in Gruppen, deren Größe dem Gauß'schen Gesetz entspricht.

Dann bilden die Zahlen der abändernden Ursachen, die auf jedes Objekt fallen, eine arithmetische Reihe; bei jeder neuen Gruppe tritt eine gleiche Zahl von abändernden Ursachen hinzu, wenn diese rein zufällig verteilt sind. Ihre Wirkung läßt sich aus dem Unterschied ermessen, den die Objekte am Beginn und Ende jeder Gruppe aufweisen. Man kann so den Gang der Entwicklung verfolgen, den ein Objekt verlaufen würde, wenn die abändernden Ursachen in jedem Stadium der Entwicklung gleich zahlreich wären; man schaltet die störende Wirkung aus, die daraus hervorgeht, daß von einer mittleren Zahl abändernder Ursachen größere Gruppen von Objekten betroffen sind, als von einer kleinen oder größeren Zahl.

Nachträgliche Auslese

Endlich kann die Häufigkeit der Grade der Abweichung durch eine nachträgliche Auslese beeinflußt werden. Erweisen sich bestimmte Gruppen als nicht lebens- oder fortbestandsfähig, so wird ihre Zahl vermindert, die der anderen Gruppen also verhältnismäßig erhöht. Eine solche Auslese wird bei der

Züchtung von Haustierrassen absichtlich herbeigeführt, indem nur diejenigen Individuen zur Aufzucht verwendet werden, die die erwünschten Abänderungen zeigen. Dies Beispiel zeigt zugleich durch die bei der Tierzucht ebenfalls angewandten Kreuzungen von Rassen und Kombinationen wünschenswertr Eigenschaften innerhalb der Rasse, wie das Moment der Auslese mit dem Moment der Mischung zusammenwirken kann, um die dem Gauß'schen Gesetz entsprechende gleichmäßige Häufigkeit aller Möglichkeiten zu beeinflussen.

In ähnlicher Weise vollzieht sich durch bloße Auslese eine Umgestaltung der wirtschaftlichen und staatlichen Betriebs- und Organisationsformen schon da, wo lediglich das Herkommen maßgebend ist. Die erfolglosen werden ausgeschieden und durch Ausdehnung und Nachahmung der erfolgreichen ersetzt.

Im gesellschaftlichen und staatlichen Leben sind ferner die Fälle häufig, in denen eine Änderung eines Zustandes eine weitere Änderung in derselben Richtung erleichtert oder auch erschwert.

Wir können diese Wirkung zwar selten messen, aber doch oft ihr Vorhandensein aus der Analyse des psychologischen und wirtschaftlichen Zusammenhangs feststellen.

So wirken in der Wirklichkeit die verschiedenen Ursachen, die die zufällige Gruppierung der Abänderungen modifizieren, gleichzeitig, verstärken und hemmen einander, ohne daß wir das Maß ihrer Wirkuno a priori bestimmen könnten. Wir können ihrer Erkenntnis nur dadurch näher kommen, daß wir uns ihre Wirkung isoliert vorstellen und mit der wirklichen Gruppierung der Dinge vergleichen. Zu einem solchen Vergleich von Zahlenreihen eignet sich am besten die graphische Darstellung, die die wirklich beobachteten Zahlen und die Gestalt der hypothetischen Funktion gegenüberstellen kann.

Graphische Darstellung

Zunächst müssen wir uns darüber klar werden, ob wir gleichen absoluten oder gleichen relativen Unterschieden der Grade des Merkmals einstweilen gleiche Bedeutung beilegen und demgemäß die Kurve der Zahlen oder der Logarithmen zeichnen können. Werden die kleinsten Werte des Merkmals negativ, so können wir deren Logarithmen nicht zeichnen; nähern sie sich dagegen der 0, ohne sie zu erreichen, so sind die Logarithmen anwendbar und notwendig, um die Unterschiede der kleinsten Werte erkennbar zu machen. Haben wir keinen Anhaltspunkt für die Entscheidung dieser Frage, so wählen wir fünf Werte versuchsweise aus, die etwa den Quartilen entsprechen und zeichnen eine Skizze der Merkmalskurve. Zeigt diese eine nahezu gleichmäßige Verteilung ihrer Steilheit auf die erste und zweite Hälfte, so kann man sich mit gleicher Bedeutung der absoluten Unterschiede begnügen. Fällt dagegen die Merkmalskurve zuerst steil ab, um dann flach auszulaufen, so zeigt dies, daß die relativen Unterschiede der Betrachtung zugrunde zu legen sind. Es ist daher dann die Merkmals- und die Häufigkeitskurve der Logarithmen zu zeichnen; ebenso wenn sich aus der Natur der Dinge ergibt, daß relative Unterschiede im Verhältnis zum Gesamtmaß der Eigenschaft so gering sind, daß die Logarithmen den Zahlen fast proportional sind.

Hat man so Kurven erhalten, bei denen nicht die Verschiedenheit der absoluten Unterschiede jede Ähnlichkeit mit der Fehlerkurve von vorneherein ausschließt, so kann man zunächst aus der Gestalt der Kurve einen Schluß darauf ziehen, ob die beobachtete Masse einheitlich und vollständig zu sein scheint. Zeigen sich zwei Gipfel der Häufigkeitskurve und dementsprechend zwei flachere Stellen der Merkmalskurve, so ist darauf zu schließen, daß die

beobachteten Gegenstände oder Erscheinungen zwei verschiedenen Typen angehören können, oder daß durch eine Auslese oder durch die Abgrenzung des Beobachtungsfeldes die mittleren Gruppen geschwächt worden sind oder auf Verminderung einzelner Gruppen durch zu enge Abgrenzung des Beobachtungsfeldes. Es tritt dann ein ähnlicher Vorgang ein, als wenn der Altersaufbau einer Stadt infolge der Binnenwanderungen, der Anwesenheit zahlreicher Soldaten, Dienstboten, jugendlicher Arbeiter usw. eine von dem Ergebnis der natürlichen Bevölkerungsbewegung abweichende Gestalt hat. Ebenso wie für das ganze Land der natürliche Altersaufbau weit weniger durch Wanderungen gestört zutage tritt, so kann man auch unter Umständen die Merkmale und Häufigkeitskurven der Fehlerkurve dadurch ähnlicher machen, daß man das Beobachtungsgebiet erweitert. Umgekehrt: wie man den Altersaufbau durch Ausscheidung der Militärbevölkerung, soweit sie den Durchschnitt übersteigt, und wenn möglich auch an den Bevölkerungsklassen, von denen nur ein bestimmtes Alter und Geschlecht an dem Orte anwesend ist, dem normalen Altersaufbau nähern kann, so kann man versuchen, durch Zerlegung der Gesamtmenge der Objekte in Teilgruppen von gleichartiger Beschaffenheit zu Häufigkeitskurven zu gelangen, die der Fehlerkurve ähnlicher sehen; man hat so in den Schwankungen der Fehlerkurve einen Hinweis auf eine Zusammensetzung der Objekte aus zwei oder mehreren Typen.

Dies ist allerdings nur ein Hinweis auf eine Möglichkeit, der erst durch den Versuch der Zerlegung oder der Erweiterung des Beobachtungsgebietes bestätigt werden müßte. Gelingt ein solcher Versuch, so hat man Teilmassen bezw. umfassendere Gesamtmassen, die wieder mit dem Fehlergesetz verglichen werden können; mißlingen die Versuche anderer Gruppierung und Umgrenzung, so ist damit immer noch nicht erwiesen, daß eine einheitliche Masse besteht, aber doch die Wahrscheinlichkeit näher gerückt, daß die Abweichungen der Häufigkeitskurve von der Fehlerkurve eine andere Ursache haben, als die Existenz verschiedener Typen oder die ungeeignete Begrenzung der Beobachtungsmasse. Es kann z. B. sein, daß die mittleren Gruppen tatsächlich durch äußere Einflüsse, durch eine Auslese vermindert sind, ihre Ergänzung zum vollen Bestand aber aus praktischen Gründen nicht möglich ist, ähnlich wie sich der Einfluß überseeischer Auswanderung nur schwer durch Zusammenzählung der Aus- und Einwanderungsländer wieder ausgleichen läßt.

Ob neben der Möglichkeit einer Auslese noch diejenige einer gegenseitigen Annäherung der Grade des Merkmals besteht, ähnlich wie bei Mischung von Flüssigkeiten, bei Kreuzung von Rassen, bei Berechnung von Durchschnitten und Beziehungszahlen, wird sich im einzelnen Fall feststellen lassen. Scheiden diese Möglichkeiten aus, ist auch eine Auslese nicht wahrscheinlich, so bleibt als Ursache der Abweichung von der Fehlerkurve nur ein verschiedener Erfolg der abändernden Einflüsse, je nach dem Grade, den das Merkmal erreicht hat.

Durch Ausscheidung der binomischen Anhäufung reduzierte Merkmalskurve

Um diesen Erfolg zu veranschaulichen, zeichnen wir zunächst die kumulative Fehlerkurve (s. Fig. 2) und teilen deren Ordinatenachse in gleiche Teile, 100 Abschnitte. Diese entsprechen dann den perzentilen Graden der Abänderung nach dem Gaußschen Gesetz. Ziehen wir durch die Teilpunkte Parallelen zur Abszissenachse bis zur kumulativen Fehlerkurve und fällen von den Schnittpunkten senkrechte Linien auf die Abszissenachse, so haben wir in der Dichtigkeit dieser Linien ein Abbild der Häufigkeit, mit der die Abweichungen vertreten sind. Errichten wir in denselben Abständen Ordinaten auf einer Parallele zur

Abszissenachse und machen wir diese dem dazugehörigen Werte der Abweichung t vom mittleren Wert proportional, so muß eine gerade Linie entstehen; denn auch die Abszissen drücken diesen Wert aus, sind also den neuen Ordinaten proportional bezw. gleich. Wir tragen nun auf denselben Ordinaten die Logarithmen der perzentilen Grade des Merkmals der statistischen Reihe auf, die wir untersuchen wollen, so sehen wir, wie weit und in welcher Richtung diese Reihe von der nach dem Gauß'schen Gesetz zu erwartenden Reihe abweicht.

Verfügen wir nicht über perzentile Grade oder andere gleiche Gruppen, so teilen wir die Ordinatenachse der kumulativen Fehlerkurve in dem Verhältnis der gegebenen Intervallgruppen, ziehen die Abszissen bis zur Fehlerkurve, dann senkrecht dazu die Linien, auf deren Verlängerung die Ordinaten abzutragen sind. Auch dann zeigt sich, inwiefern die Wirkung der abändernden Ursachen je nach dem erreichten Grade des Merkmals verschieden ist; je größer sie ist, desto steiler verläuft an der betreffenden Höhe die Kurve.

Wird die Wirkung der abändernden Ursachen in dem Maße gehemmt, als sie den Wert des Merkmals von der Mitte entfernt, so wird die auf das Fehlergesetz reduzierte Merkmalskurve die entgegengesetzte Form der nicht reduzierten logarithmischen Merkmalskurve zeigen; sie wird an den Enden flach, in der Mitte steiler verlaufen. Das gleiche tritt dann ein, wenn Objekte, die sich weit von dem Typus entfernen, wenn nicht mehr zu derselben Gattung gerechnet werden, oder aus einem anderen Grund nicht in die Statistik aufgenommen worden sind. Wird die Wirkung der abändernden Ursachen nur gegen das Maximum gehemmt, so wird nur der obere Teil der Kurve flacher. Ist die Wirkung der die Werte des Merkmals erhöhenden Ursachen allgemein umsomehr gehemmt, je größer diese Werte jeweils schon sind, so wird die logarithmische reduzierte Merkmalskurve eine stetige Krümmung nach abwärts aufweisen.

Ersetzt man die Logarithmen wieder durch die dazu gehörigen Zahlen, so nimmt die letzte der Kurven wieder die doppelt gebogene Form mit größter Steilheit in der Mitte an, die wir bei den Kurven gehemmter geometrischer Progression beobachtet haben. Ebenso werden bei den übrigen Kurven die absoluten Unterschiede dargestellt. Die Hemmungen zeigen sich auch bei diesen Kurven noch in der Abweichung von dem Verlauf, den eine Kurve geometrischer Progression sonst nehmen würde, sind aber viel schwerer zu erkennen, als bei der logarithmischen reduzierten Merkmalskurve, die einfach mit einer graden Linie zu vergleichen ist. Schwache Hemmungen können von der allgemeinen Tendenz der Aufwärtskrümmung verwischt werden und gegen das Minimum verläuft die nicht logarithmische Kurve bei großer Gesamtabweichung so flach, daß überhaupt keine feineren Unterschiede mehr zu bemerken sind. Nur bei geringer Gesamtabweichung ist die logarithmische Kurve entbehrlich, die nicht logarithmische Kurve sollte aber immer außerdem ausgeführt werden, um die tatsächliche Bedeutung der dargestellten Zahlen zu veranschaulichen.

Kann das Merkmal auch negative Werte annehmen, so ist eine logarithmische Behandlung nicht ohne weiteres möglich. Zunächst ist die arithmetische Merkmalskurve auf das Fehlergesetz reduziert zu zeichnen. Aus ihrem Charakter ergibt sich, ob wirklich die Wirkungen der abändernden Ursachen eine arithmetische Reihe bilden, oder ob sie die Form einer geometrisch progressiven Kurve zeigt, die durch Subtraktion einer konstanten Größe zum Teil negativ geworden ist. Aus der Kurve ergibt sich dann diese negative Größe und nach deren Eliminierung kann man die Logarithmen der Reihe aufsuchen und deren Kurve mit einer geraden Linie vergleichen.

Tragweite der Schlüsse aus der Gestalt der reduzierten Merkmalskurve

Wie schon erwähnt, kann aus der Abweichung der reduzierten Merkmalskurve nur dann auf eine Verschiedenheit der Wirkung der abändernden Ursachen mit Sicherheit geschlossen werden, wenn es unmöglich ist, daß die Reihe durch andere Umstände, z. B. natürliche oder künstliche Auslese, Vereinigung und gegenseitige Annäherung mehrerer Typen beeinflußt worden ist. Aber auch wenn dies nicht von vorneherein feststeht, kann die Reduktion der Merkmalskurve über das Maß der Abweichung und ihre Gruppierung Aufschlüsse geben, ja diese unter Umständen erst deutlich erkennen lassen.

Sind z. B. zwei Typen zusammengefaßt, deren Gipfel in der Häufigkeitskurve (s. Fig. 6) so nahe zusammenliegen, daß sie einen einzigen breiteren Gipfel bilden, so wird die reduzierte Merkmalskurve in der Mitte eine flachere Stelle zeigen. Diese ist leichter zu erkennen, als die Verbreiterung der Häufigkeitskurve, und kann Veranlassung geben, zu untersuchen, ob durch Teilung der Masse in zwei Typen sich regelmäßigere Linien erzielen lassen, oder ob die mittleren Gruppen durch eine Besonderheit des Beobachtungsmaterials geschwächt sind. In ähnlicher Weise regen flache Stellen an den Enden der reduzierten Kurve zu der Untersuchung an, ob die Extreme vielleicht von der Statistik nicht umfaßt sind, ob sie durch die natürliche oder soziale Auslese beseitigt sind, oder ob die Wirkung der abändernden Ursachen gehemmt wurde. Die reduzierte Kurve kann ebensowenig diese Frage beantworten, wie der Vergleich der Häufigkeitskurve mit der Fehlerkurve durch Schätzung nach dem Augenmaß; sie gibt nur der Schätzung eine größere Leichtigkeit und eine Genauigkeit, die sonst nur durch Rechnung zu erzielen wäre. Vor der Rechnung besitzt sie den Vorzug größerer Anschaulichkeit, Übersichtlichkeit und Einfachheit.

Anhang

Rechnerische Ausgleichung

Es ist bereits ausgeführt worden, wie man die Kurven konstruktiv ausgleichen kann. Oft empfiehlt es sich aber, die Ausgleichung rechnerisch durchzuführen, sei es wegen der größeren Genauigkeit, sei es, weil man die zahlenmäßigen Resultate zu haben wünscht.

Bei der rechnerischen Durchführung ist man nicht auf die angegebene Methode[1]) beschränkt, diese empfiehlt sich aber m. E. für den Fall, daß die Schwankungen nicht periodisch sind.

Die rechnerische Ausgleichung ist nur dann leicht durchzuführen, wenn die Intervalle gleich sind. Unter Umständen läßt sich die Gleichheit dadurch herstellen, daß Intervalle mit der Häufigkeit = 0 eingeschoben werden; andernfalls ist die Reihe in eine solche mit gleichen Intervallen umzugestalten. Zu diesem Zweck nimmt man an, daß innerhalb der gegebenen Intervalle die Häufigkeit gleichmäßig ist. Löst man die Intervalle in Teile auf, die sich zu neuen gleichen Intervallen zusammenfügen lassen, so kann man die Zahl der Objekte für diese berechnen, indem man den Umfang des Teilintervalls mit der Häufigkeit multipliziert, die Produkte addiert und durch den Umfang des neuen gleichen Intervalls dividiert. Der Erfolg ist derselbe, wie bei Verwandlung der Rechtecke, oder bei Verwendung der perzentilen Grade der polygonförmigen Merkmalskurve.

[1]) Abgesehen von den Methoden mit Hilfe höherer mathematischer Operationen.

Für periodische Schwankungen

Ist die Schwankung der Häufigkeitskurve periodisch, sind z. B. die auf 0 oder 5 endigenden Messungen stärker, die benachbarten schwächer vertreten, so wird man Gruppen der Messungswerte von der Länge der Schwankungsperiode bilden. Diese enthalten dann stets ein Maximum und ein Minimum. Anstelle jeder der angegebenen Häufigkeiten setzt man den Durchschnitt der Gruppe, deren Mitte sie bildet, z. B.

für Intervall 3 die Gruppe 1—5
für Intervall 4 die Gruppe 2—6

Bei gerader Zahl der Glieder wird der Durchschnitt zwischen die Intervalle gesetzt, z. B. der Durchschnitt für die Werte 0—9 zwischen 4 und 5 als 4,5 1—10 zwischen 5 und 6 als 5,5.

Ist die Reihe nicht periodisch, so würde die Zahl der zusammenzufassenden Intervalle nur willkürlich bestimmt werden können. Die Durchschnittsbildung würde die Oszillationen zwar vermindern, aber doch bestehen lassen. Denn diese Gruppen von Intervallen unterscheiden sich nur durch die Endglieder. Ist die Reihe periodisch, so reicht die Gruppe von einem Maximum bis nahe an das nächste; die nächste Gruppe enthält das erste Maximum nicht mehr, dafür aber das zweite; die dritte vertauscht eins der dem Maximum benachbarten Glieder gegen eins in netsprechender Stellung. So wird die Reihe der Durchschnitte von der periodischen Oszillation nicht berührt. Bei nicht periodischer Oszillation kann es aber vorkommen, daß der nächste Durchschnitt anstelle eines Maximums ein Minimum enthält. Dazu kommt, daß die Hebungen und Senkungen nicht an der Stelle auftreten (vergl. Benini,[1]) Kaufmann), wo sie bei der ursprünglichen Reihe liegen, sondern eine halbe Länge der Ausgleichungsgruppen weiter. Die Reihe wird also nicht nur nicht ausgeglichen, sondern ihre Oszillationen zeigen auch keine Ähnlichkeit mehr mit denen der ursprünglichen Reihe. Um diese Ähnlichkeit zu bewahren, ist es nötig, die mittleren Glieder in den Durchschnitt stärker einzusetzen; und zwar in absteigender Reihe bis zum Ende. Eine Erhöhung der ursprünglichen Reihe bewirkt dann nicht ein plötzliches Ansteigen der Reihe der Durchschnitte, sondern eine allmähliche Hebung und Senkung.

Diese Wirkung kann man erzielen, wenn man die Durchschnittsbildung wiederholt. Hat man z. B. je fünf Intervalle zusammengefaßt, so wird in der Reihe a, b, e, d, e, f, g, ...

das Glied c durch $\frac{a+b+c+d+e}{5}$

„ „ d „ $\frac{b+c+d+e+f}{5}$

ersetzt usw.

Wiederholt man diese Durchschnittsbildung, so tritt anstelle des Gliedes e der Durchschnitt

$$\frac{\frac{a+b+c+d+}{5} + \frac{b+c+d+e+f}{5} + \ldots}{5}$$

$$= \frac{a+2b+3c+4d+5e+4f+3g+2h+i}{25}$$

Bei Zusammenfassung von je drei Gliedern hat man

$$c = \frac{a+2b+3c+2d+e}{9}$$

[1]) Benini, Principi di statistica metodologien. S. 129.

allgemein gesprochen bei zweimaliger Zusammenfassung von je n Gliedern hat man den Durchschnitt aus 2n—1 Gliedern mit den Koëffizienten 1, 2, 3 . . . n, (n—1), . . . 2, 1 und dem Nenner 2^n

Diese Formel bezeichnet zugleich den Bruchteil der von jedem Glied der Reihe auf die benachbarten Stellen verteilt wird, um zur Bildung der Durchschnitte beizutragen. Graphisch entspricht dieser Art der Ausgleichung die Ersetzung der Messungshäufigkeiten durch Dreiecke von mehrfacher Breite eines Intervalls und entsprechend verminderter Höhe. Zur Erleichterung der Rechnung wird man darnach streben, als Nenner 10 zu erhalten; zu diesem Zweck kann man dem Durchschnitt

$$\frac{b + 2c + 4d + 2e + f}{10} \text{ oder}$$

$$\frac{\frac{1}{2}a + b + 2c + 3d + 2e + f + \frac{1}{2}g}{10}$$

für d setzen, oder bei starker Ausgleichung

$$\frac{a + 2b + 3c + \ldots\ldots + 10k + 9l + \ldots\ldots + 2t + n}{100}$$

für k als mittelstes Glied der Reihe a bis u.

Diese Art der Ausgleichung entspricht nun aber nicht vollkommen dem Gedanken, daß die wirkliche Reihe nur ein zufälliger Fall der möglichen Reihen ist, deren Durchschnitt als idealer Typus dargestellt wird. Sie nähert sich dem Ideal insofern, als sie als wahrscheinlicher voraussetzt, daß die Abweichungen klein als daß sie groß sind, nimmt aber ein willkürliches Verhältnis ihrer Häufigkeit zu ihrer Größe an, das durch arithmetische Reihen ausdrückbar ist.[1])

A n m e r k u n g: Nennt man die Abweichung x, die Wahrscheinlichkeit ihres Vorkommens y, so ist angenommen, daß

$$y = a - b\,x$$

ist, wobei a und b willkürliche Konstanten sind. Dies ist die Gleichung einer geraden und wenn x stets positiv genommen wird, der Durchschnitt eines Dreiecks.

Auch die Annahme, daß die Wahrscheinlichkeit der Abweichung im umgekehrten Verhältnis zu ihrer Größe steht, befriedigt nicht, weil dann die Wahrscheinlichkeit für die Abweichung o unendlich viel größer ist, als die jeder meßbaren, also nicht unendlich kleinen Abweichung.[2]) Dies würde dazu führen,

[2]) Anstatt eines Dreiecks wird jede Gruppe zu einer von Hyperbeln begrenzten Fläche.

eine Ausgleichung überhaupt unmöglich zu machen; bei nicht absolut genauen Messungen ist aber zu vermuten, daß die Wahrscheinlichkeit einer Abweichung innerhalb der Grenzen der Genauigkeit der Messung annähernd gleich groß ist, und dann erst etwa in dem umgekehrten Verhältnis an ihrer Größe abnimmt.

Zu einem besseren Resultat gelangt man, wenn man mehrfache Durchschnittsbildung von je zwei benachbarten Häufigkeiten vornimmt.

A n m e r k u n g: Man verteilt so jede Häufigkeit in der Weise über die benachbarten Intervalle, wie es dem Gesetz zufälliger Abweichungen entspricht. Dies geht von dem Gedanken aus, daß sich mehrfache kleine Abweichungen teils summieren, teils kompensieren. Die Wahrscheinlichkeit der einzelnen Ergebnisse wird durch die Koeffizienten einer binomischen Entwicklung dividiert durch deren Summe ausgedrückt; die des Mittelgliedes ist $\frac{1}{2^n} \frac{n!}{\left(\frac{n}{2}\right)!}$ die Wahrscheinlichkeit für ein Glied mit der xfachen Abweichung ist

$$\frac{1}{2n} \cdot \frac{n!}{\left(\frac{nx}{2} + x\right)!\left(\frac{n}{2} - x\right)!}$$ wobei $\left(\frac{n}{2} + x\right)$ und $\left(\frac{n}{2} - x\right)$ ganze Zahlen sind.

Umfaßt das Intervall a Abweichungen, so ist die wahrscheinliche Häufigkeit für

$$\frac{a}{2n} \cdot \frac{n!}{\left(\frac{n}{2} + x\right)!\left(\frac{n}{2} - x\right)!}$$

Vergleiche die Ausführung über das Gauß'sche Gesetz und seine graphische Anwendung.

Man braucht dann nicht die Zahl zu einem Durchschnitt zu vereinigenden Glieder willkürlich zu bestimmen, sondern kann die Durchschnittsbildung mit der geringst möglichen Abweichung von der Urtabelle beginnen, und so lange fortsetzen, bis die Ausgleichung genügt. Die Abweichung beträgt jedesmal nur ½ Intervall, also nur so viel, als die schon in der Tabelle liegende Ungenauigkeit. Geometrisch ersetzt man das Rechteck für jede Häufigkeit durch eine Figur, die zuerst langsam, dann stärker ansteigt, in der Mitte aber abgerundet ist. Leichter ist es jedoch, die Mittelpunkte der Verbindungslinien fortgesetzt zu verbinden. Rechnerisch wird das Glied d zuerst

$$\frac{\frac{e + d}{2} + \frac{d + e}{2}}{2} = \frac{c + 2d + e}{4}$$

sodann

$$\frac{\frac{b + 2c + d}{4} + \frac{c + 2d + e}{2} + \frac{d + 2e + f}{4}}{4}$$

$$= \frac{b + 4c + 6d + 4c + f}{16} \text{ u.s.w.}$$

Die Ausgleichung von Reihen wird nun oft in der Weise ausgeführt, daß nur diejenigen Glieder wiedergegeben werden, für welche sich ein Durchschnitt in dieser Weise ermitteln. Die Endglieder fallen dadurch fort. Dies ist korrekter für Reihen, die nur Ausschnitte aus einer größeren Gesamtreihe sind, z. B. aus zeitlichen Entwicklungen. Man kann dann nicht wissen, wie der Durchschnitt sein würde, wenn zu seiner Bildung die benachbarten Glieder mit herangezogen werden würden.

Bei Reihen von Messungen und Häufigkeiten ist aber die ganze Reihe gegeben, und bekannt, daß die folgenden Glieder den Wert 0 haben. Man kann daher diesen Wert so oft einsetzen, als erforderlich ist. Die Summe der Objekte bleibt dann dieselbe wie vorher. Allerdings wird die Zahl der Intervalle bei jeder Ausgleichung nach jeder Seite um ½ vergrößert. Es ist deshalb in Fällen eines natürlichen Maximums und Minimums der Grade des Merkmals ebenso wie bei der graphischen Ausgleichung eine Modifikation anzubringen. Wenn sich aus der Natur der Eigenschaft ergibt, daß sie gewisse Grenzen nicht überschreiten, z. B. nicht unter 0 sinken oder nicht über 100 steigen, so würde es ebenso falsch sein, die unter dem Minimum bezw. über dem Minimum liegenden Glieder anzuführen, als sie fortzulassen. Sie sind vielmehr mit den benachbarten zu vereinigen, die Reihe ist an der tatsächlichen Grenze umzukehren und zu addieren. Dies ergibt sich aus folgender Erwägung: damit die Gesamtsumme erhalten bleibt, müssen die Glieder der ursprünglichen Reihe in den Zählern der Durchschnitte zusammen so oft enthalten sein, als der Nenner beträgt. Kann man sich die Reihe nicht durch Glieder vom Werte 0 verlängert denken, so fehlen in den Zählern Elemente; in diese Lücken sind die Glieder

einzusetzen, die sonst für Durchschnitte jenseits der Grenze verwendet worden wären. Hat man z. B. die Reihe

$$a, b, c, d \ldots .$$

wobei a ein Minimum ist, so bildet man die Durchschnitte

$$a,\quad \frac{a+b}{2},\quad \frac{a+c}{2},\quad \frac{c+D}{2},\ \ldots\ldots\ldots$$

$$\frac{3a+b}{4},\quad \frac{a+2b+c}{4},\quad \frac{b+2c+d}{4},\ \ldots\ldots$$

$$\frac{3a+b}{4},\quad \frac{4a+3b+c}{8},\quad \frac{a+3b+3c+d}{8}.\ \ldots\ldots$$

$$\frac{10a+5e+c}{16},\quad \frac{5a+6b+4c+d}{16},\quad \frac{a+4b+6c+4d+4}{16}$$

Die Grenzglieder bleiben bei der 1., 3., 5. Durchschnittsbildung unverändert und treten bei den 2., 4., 5. wieder ein. Die Reihe hat dann ihre Länge wieder erhalten. Das Minimum bezw. Maximum wirkt so in ähnlicher Weise, wie eine feste Wand, die eine zähe Flüssigkeit begrenzt.

Dieselben Ausgleichungsmethoden kann man auf die Zahlen der Objekte anwenden, die bei gleichen Intervallen in den kumulativen Reihen angegeben sind. Das Ergebnis ist dann dasselbe; es ist aber leichter, die Zahlen der einzelnen Intervalle auszugleichen, weil diese kleiner sind, und aus dem Ergebnis die kumulative Reihe neu zu bilden. Auch graphisch ist die Ausgleichung der einfachen Häufigkeitskurve leichter, weil die Verbindungslinien der Ordinaten kürzer sind.

Alle diese Methoden bezwecken die Wiederherstellung der wirklichen, durch die Ungenauigkeit der Messung und durch die Zusammenfassung in Gruppen zum Zweck der Tabellenbildung deformierten Reihe. Sie können ihr Ziel nur unvollkommen erreichen, bewirken aber doch, wie die Figuren durch Vergleich mit der durchsichtigen Kopie der ursprünglichen Reihe zeigen, daß sich das Diagramm der Wirklichkeit wieder mehr nähert, als es die bloße Wiedergabe der Zahlen einer zusammengefaßten Tabelle durch eine Reihe von Rechtecken vermöchte. Mit Hilfe höherer Mathematik sind noch vollkommenere Methoden erdacht worden, um die Wirklichkeit wieder herzustellen und aus dem beobachteten Spezialfall auf das vermutlich zugrunde liegende allgemeine Gesetz zu schließen. Ihre Anwendung setzt aber schwierige und umständliche Berechnungen voraus, die sowohl die Herstellung verteuern und verzögern, als auch die Nachprüfung erschweren oder unmöglich machen würden. Ihre Anwendung in den statistischen Publikationen ist daher weder allgemein zu erwarten, noch dürfte sie sich m. E. allgemein empfehlen.

Dagegen scheint mir, daß die statistischen Ämter mit verhältnismäßig geringerer Mühe der Wirklichkeit noch näher kommen könnten, wenn sie die Diagramme nicht erst aus zusammengefaßten Tabellen, sondern stets aus dem so wenig als möglich gekürzten Urmaterial herstellen ließen. Das Urdiagramm würde dann allerdings erheblichen Umfang annehmen, ließe sich aber durch photographische Verkleinerung auf jedes gewünschte Format reduzieren.

Da bei den Merkmals- oder Kumulations-Kurven eine Ausgleichung zur Beseitigung der Oszillation nicht notwendig ist, kann die direkt gefundene Linie ohne weiteres verwendet werden. Es wäre also zur Herstellung des Diagramms

keine wissenschaftliche, sondern nur technische Vorbildung erforderlich. Vielleicht würden im Bedarfsfalle auch Apparate erfunden werden, die durch die Auszählung zugleich mechanisch ein Diagramm herstellen; im Prinzip scheint die Schwierigkeit nicht groß zu sein, da es sich nur um feine Bewegungen eines zeichnenden Stiftes handelt, der von einer Maschine mit größter Präzision und ebenso leicht bewirkt werden könnte, als die Drehung eines Zählapparates. In der Technik sind Diagramme zeichnende Meßapparate bereits vielfach im Gebrauch.



C. Diagramme der Einkommen.

Die graphischen Darstellungen können durch die gleichzeitige Veranschaulichung mannigfaltiger Zahlen und Zahlenbeziehungen dazu dienen, Gesetzmäßigkeiten der Entwicklung und Zusammenhänge verschiedener Erscheinungen aufzufinden.

So läßt eine Kurve zeitlicher Entwicklung erkennen, ob die Zunahme progressiv ist, also sich noch steigert, oder ob sie nachläßt, so daß ein Aufhören der Zunahme, vielleicht ein Rückgang zu erwarten steht, oder ob die Entwicklung periodisch auf- und abschwankt.

Der Vergleich von mehreren Kurven zeitlicher Entwicklung zeigt deutlicher, als der Vergleich von Zahlenreihen, ob ein Zusammenhang zwischen den Vorgängen in den Zahlen zum Ausdruck kommt oder nicht.

Aus diesem ersten Eindruck der Kurven läßt sich eine vorläufige Hypothese gewinnen, indem man die Haupttendenz der Größenbeziehungen auf einen einfachen Ausdruck bringt, z. B. daß die eine Reihe der andern proportional oder entgegengesetzt proportional verläuft, daß periodische Schwankungen auftreten usw. Zur Prüfung dieser, aus der Statistik gewonnenen Hypothesen sowie solcher, die sich aus der Natur der Dinge ergeben, läßt sich wieder die graphische Darstellung mit Vorteil anwenden.

Zu diesem Zweck kann man eine Figur zeichnen, die den Bedingungen der Hypothese entspricht und sie mit der statistisch gewonnenen Figur vergleichen, z. B. man vermutet, daß die Bevölkerung sich so vermehrt, wie ein auf Zinseszins gelegtes Kapital, weil die Zunahme von der jeweils erreichten Größe abhängig ist, und zeichnet eine Kurve in geometrischer Progression, oder man vermutet, daß die Häufigkeiten der Abweichungen in einer sachlich quantitativen Reihe sich so verhalten, wie die Häufigkeiten zufälliger Beobachtungsfehler nach dem Gauß'schen Gesetz und zeichnet zum Vergleich eine binonische Kurve.

Eine solche Prüfung ist dann am leichtesten durch den bloßen Augenschein anzuführen, wenn den Bedingungen der Hypothese durch eine gerade Linie genügt wird. Ist dies nicht der Fall, so entsteht die Schwierigkeit, die Kurve mit einer krummen Linie zu vergleichen; noch schwieriger ist es, einer Reihe von krummen Linien, die der Hypothese genügen würden, diejenige erst auszusuchen, die der statistischen Kurve am nächsten kommt.

In diesen Fällen wird die Aufgabe erleichtert, wenn man aus den ursprünglichen Zahlen eine abgeleitete Reihe bildet, die bei Zutreffen der Hypothese durch ein gerade Linie dargestellt werden könnte. Aus dem Abweichen dieser abgleiteten Kurve zeigt sich, inwiefern die Hypothese sich mit der Wirklichkeit noch nicht deckt, und daher modifiziert werden muß, bis eine genügende Übereinstimmung erzielt ist.

Im folgenden sollen nach diesen Gesichtspunkten die Diagramme der Einkommensverteilung und die Hypothesen besprochen werden, die zur Erklärung der dabei zutage tretenden Regelmäßigkeiten aufgestellt werden können.

Die einfachste Form der Darstellung ist die bekannte Frequenz- oder Häufigkeitskurve, bei der die eine Achse eines rechtwinkeligen Koordinatsystems die Grade des Merkmals, also hier den Betrag des Einkommens, die andere Achse das Verhältnis der auf eine Abstufung, z. B. 1200—1500 Mk., entfallenden Personen zu dem Umfang der Abstufung darstellt. Ein solches

Diagramm hat z. B. Otto Ammon[1]) auf Grund der sächsischen Einkommen-

[1]) O. Ammon: Die Gesellschaftsordnung und ihre natürlichen Grundlagen. 3. Aufl. Jena 1900. S. 100.

steuerstatistik gezeichnet und der Galton'schen Kurve der Begabungen gegenübergestellt.

In der Tat läßt gerade die sächsische Einkommensteuerstatistik, die alle erwerbstätigen Personen umfaßt, und die Einkommen der Individuen, nicht der Familien berücksichtigt, eine gewisse Ähnlichkeit zwischen der Einkommenskurve und der symmetrischen binonischen Kurve erkennen, die nach Galton[2])

[2]) Zitiert S. 61 bei Ammon.

ein Abbild der Verteilung der Begabungen geben soll. Insbesondere ist ersichtlich, daß die Häufigkeit mit fallendem Einkommen zunimmt, und nur die unterste Einkommensstufe wieder weniger zahlreich besetzt ist, als diejenige, die noch eine gordnete Lebenshaltung ermöglicht; d. h. nach Ammon S. 218 über 400 Mk. Das Maximum befindet sich etwa bei 6000 Mk. Eine schmal auslaufende Spitze erstreckt sich unter das Existenzminimum (100 Mk.), ja sogar unter die Grenze der positiven Produktivität und umfaßt Personen mit „negativem Einkommen" (S. 103). Auf der oberen Seite läuft die Einkommenskurve Ammons außerordentlich weit aus, so daß eine graphische Darstellung praktisch unmöglich wird. Ein Einkommen von 500 000 Mk. würde erst auf einem Format von 8 m Länge hergestellt werden können, da je 500 Mk. Differenz 8 mm erfordern. So erstreckt sich die Figur nur bis 3300 Mk., die wiedergegebene Tabelle bis 9600 Mk. und hat noch die Stufen 1600, 800, 500, 300, 0 Mk., die die Bevölkerung in acht Teile zerlegen. Trotz der geringen Detaillierung gibt Ammon durch Interpolation eine sehr anschauliche Kurve (Fig. 1) die durch Vergleichung der Prozentzahlen für 1879, 1890 und 1898 auch die zeitliche Entwicklung erkennen läßt.

Von der Galton'schen Kurve der Begabungen und der binonischen Kurve unterscheidet sie sich insofern, daß Einkommen zwischen 800 und 1600 Mk. weniger häufig, solche über 1600 Mk. häufiger sind und nicht von etwa 2000 Mk. an aufwärts verschwinden, sondern sich anscheinend unbegrenzt, wenn auch mit sehr geringer Häufigkeit noch über eine 1000fach größere Strecke ausdehnen.

Diese Abweichung erklärt[1]) Ammon als Wirkung sozialer und wirtschaft-

[1]) Zum Teil erklärt sich nach Ammon diese Abweichung als Wirkung des verschiedenen Beginns des erwerbsfähigen Alters bei den sozialen Klassen sowie der Einrechnung der Aktiengesellschaften. S. 102.

licher Auslese, indem ein Aufsteigen über die Mittelmäßigkeit hinaus besonders erschwert ist, eine weitere Steigerung des Einkommens jedoch wieder leichter wird. (S. 180.) So finden wir schon bei Ammon das Diagramm zur Veranschaulichung der Statistik, zum Vergleich mit der als Normalform angesehenen binomialen Kurve der Wahrscheinlichkeiten und Hypothesen, um die Abweichung zu erklären. Die Figur ist aber außer durch Ausgleichung noch nicht umgestaltet, um die Vergleichung zu erleichtern. Diese geschieht durch Augenmaß, sowie durch Berechnungen, die die ausgeglichene Kurve nicht berücksichtigen.

Einen großen Fortschritt bedeutete die Anwendung der kumulativen, logarithmischen Kurve auf die Einkommenstatistik durch Pareto.[1]) Indem er die

[1]) Pareto: a) La legge domanda Giornale degli Economisti 1895; b) Receuil proposé par la Faculté de Droit de Lausanne 1896; c) Cours d' économie politique, II, §§ 950 ff.

Logarithmen der Einkommen auf der Abszissenachse anordnete, und die Ordinaten nach den Logarithmen der Zahl der Personen bemaß, die mehr als das betreffende Einkommen bezogen, zeigten sich mit überraschender Regelmäßig-

keit nahezu gerade Linien von wenig abweichender Neigung für die verschieden Länder und Zeiten.

In der Regel schien die Gleichung[2]) einer Geraden

[2]) Zur Vermeidung von Verwechslungen erscheint eine einheitliche Bezeichnung der Größen notwendig, die mit den Gleichungen im Anhang zusammengestellt und mit den entsprechenden Gleichungen der Originalwerke verglichen sind.

$$\log N = \log H - \alpha \log R$$

$$\text{d. h. } N = \frac{H}{R^{\alpha}}$$

die Linie mit genügender Genauigkeit wiederzugeben; in einzelnen Fällen war für die Kurve aler Einkommen und stets für die der getrennten Einkommen aus Besitz und aus Arbeit die Einfügung einer oder zweier weiteren Konstanten a und b notwendig, so daß die Gleichungen

$$\log N = \log H - \alpha (\log R + a)$$

$$\text{bezw. } \log N = \log H - \alpha (\log R + a) + b R$$

entstehen.

In seiner ersten Veröffentlichung 1895 (a) berechnete Pareto anstatt α den Wert h aus der Ableitung der Gleichung (2)

$$y = \frac{-dN}{d(\log)R} = \frac{\alpha H}{R^{\alpha + 1}}$$

wobei y die Häufigkeit bedeutet, und schreibt

$$y = \frac{A}{x^{h}}$$

Später hat Pareto die Gleichungen (1) und (2) benutzt, die eine leichtere Berechnung von α ermöglichen und sich enger an die Figur anschließen.

In dieser Kumulativen beiderseits logarithmischen Kurve der Einkommen liegt m. E. vor allem die Bedeutung der Pareto'schen Entdeckung. Er hat damit ein Mittel gefunden, um in den mannigfaltigen Wirtschaftsverfassungen eine Regelmäßigkeit der Einkommensverteilung dem Beschauer unmittelbar vor Augen zu führen, die er treffend mit der Ähnlichkeit der Krystalle derselben chemischen Substanz vergleicht. Indem die Grundform durch eine gerade Linie dargestellt ist, wird zugleich möglich, auch die feineren Unterschiede zu erkennen und zu messen. So kann z. B. die zeitliche Entwicklung der Einkommen nach Größe und Verteilung durch Pareto'sche Kurven einfach und korrekt veranschaulicht und gemessen werden, wie Bresciani[1]) überzeugend nachgewiesen hat.

[1]) C. Bresciani, Über die Methoden der Einkommenverteilungsstatistik, Conrads Jahrbücher, März 1907 und Tabacovici, die Methoden der Einkommensverteilungsstatistik Leipzig 1914

Nachdem durch die „courbe des revenus" Pareto's die Bahn gebrochen war, wurden durch Andere besondere Zwecke mit größter Vollkommenheit erreicht. Corrado Gini[1]) fand seinen „indice di concentrazione", dessen Kurve sich dem wirklichen Aufbau der Einkommen noch enger anschließt und die störenden weiteren Konstanten entbehrlich macht; Benini[2]) zeigte die Anwendbarkeit des beiderseits logarithmischen Diagramms auf wirtschaftliche Massenerscheinungen verschiedenster Art, die mit der Verteilung der Einkommen oder Vermögen zusammenhängen. Während so die vollständige und genaue Darstellung

[1]) C. Gini, Indici di concentrazione e di dipenza.
[2]) R. Benini, Statistica metodologen, Torino 1906.

der Wirklichkeit mit Erfolg gefördert wurde, konnte das Problem der kausalen Erklärung noch zu keiner endgültigen Lösung gelangen.[3]) Seine Wichtigkeit wurde von Anfang an nicht nur von der Kritik betont,[4])[5]) sondern auch von Pareto[6]) selbst durch die Tat anerkannt, indem er im Anhang eine Hypothese aufstellte, die unter Anwendung der Wahrscheinlichkeitsrechnung die Einkommensverteilung auf ein in der menschlichen Natur liegendes Gesetz der Verteilung der Fähigkeiten zurückführte. Von Gini[7]) wurde ein weiteres Moment,

[3]) Hochsinger, Österr. Stat. Monatsschrift 1911, S. 632 ff.
[4]) [5]) Edgeworth, La curva della entrata e la curva della probabilità, Giorn. d. Econ. März 1897.
[6]) Pareto Cours, d' économie politique II.
[7]) Gini, Il diverso accrescimento delle classe sociale e la concentrazione della richezza, Giornale degli economisti, Roma 1909, S. 1.

die geringere natürliche Vermehrung der reicheren Klassen, hervorgehoben, das die Steigerung der Ungleichheit ceteris paribus zur Folge haben muß. Im Grunde sind alle wirtschaftlichen Tatsachen und Zusammenhänge, die auf die Einkommensverteilung von Einfluß sind, ein Teil des Kausalnexus, dessen Kenntnis die Regelmäßigkeit der Einkommensverteilung aus einem empirischen zu einem rationellen Gesetz machen würde.[1]) Aus der Fülle der näheren und entfernteren Ursachen sollen im folgenden zwei Komplexe hervorgehoben werden, die mir auf die Gestaltung der Einkommenskurve den größten Einfluß auszuüben scheinen: die Notwendigkeit, das Existenzminimum zu gewinnen und die Möglichkeit, durch Vermögensbildung das Einkommen progressiv zu steigern. Die erstere Tatsache wirkt darauf hin, daß über dem Existenzminimum eine Verdichtung auftritt, letztere, daß der obere Teil der Kurve sich qualitativ ausdehnt und quantitativ verdünnt und so das kumulative, logarithmische Diagramm sich einer geraden Linie nähert.

[1]) Vergl. Pareto Cours II § 962, s. Anhang.

Um sich gegen den Vorwurf zu rechtfertigen, daß seine Formel die Tatsachen nur beschreibe, nicht erkläre, zieht Pareto die Parallele mit den Keplerschen Gesetzen, den Plantenbahnen, deren Erklärung erst Newton gelungen sei, die aber trotzdem wertvoll seien, sowohl an sich, als deshalb, weil sie zur Newtonschen Theorie und der noch unvollendeten, genauen Erforschung die Grundlage bilden.

Nach heutiger Auffassung kann eine Formel überhaupt nie etwas erklären, sondern nur beschreiben; auch die Newton'schen Gesetze sind nur Beschreibungen. Der Wert einer Formel besteht darin, daß sie die Vorgänge in einheitlicher und verhältnismäßig einfacher Form allgemein gültig darstellt, und so die Grundlage liefert zur Erfassung der geordneten Aufeinanderfolge der Ereignisse, die wir Kausalität nennen.

Selbstverständlich gelten ale Bedenken, die gegen den Schluß aus den Steuereinschätzungen auf die wirklichen Einkommen geltend gemacht werden können, in genau demselben Maße, wenn andere nicht durch Zahlenreihen, sondern durch Diagramme ausgedrückt werden. Aber ebensowenig wie man auf den Gebrauch der Zahlen deshalb verzichten wird, weil diese an sich eine bestimmte Größe bedeuten, und Zweifel an dieser Bestimmtheit erst im Text zum Ausdruck kommen können, braucht man die graphischen Darstellungen auf diejenigen Zahlen zu beschränken, deren Richtigkeit feststeht. Im Gegenteil, sie kann zur Aufdeckung von Irrtümern führen.

Einer Zahlenreihe sieht man nicht sofort an, in welcher Weise sie verläuft. In vielen Fällen ist nun aus der Erfahrung an ähnlichen Reihen zu erwarten oder aus der Natur der Dinge zu schließen, daß der Verlauf eine Reihe in der Regel eine gewisse Form haben wird. Z. B. verändert sich die Bevölkerung

einer Gemeinde in der Regel nur allmählich und fast kontinuierlich durch Geburten, Sterbefälle, Zu- und Abwanderung einzelner Personen. Nur bei Eingemeindungen, Verlegung von Truppenteilen, großen Betrieben, Epidemien und anderen außergewöhnlichen, meist notorischen Ereignissen tritt eine sprunghafte Änderung der Bevölkerungszahl ein. Eine solche Unstetigkeit tritt im Diagramm viel deutlicher hervor, als in der Zahlenreihe. Ebenso zeigt sich eine Abweichung von der normalen jahreszeitlichen Schwankung. Enthalten die Zahlen eine solche Abweichung oder ein Sprung in einem normalerweise stetigen Verlauf, so tritt dieser in einem Diagramm auf den ersten Blick hervor und bezeichnet die Punkte, wo die Untersuchung einzusetzen hat, ob ein erheblicher Fehler sich eingeschlichen hat oder eine besondere Ursache die Abweichung hervorbringt. Ein bekannter Fall, in dem die Unregelmäßigkeit des Verlaufs der Reihe auf Mängel der Angaben zurückgeführt werden muß, ist die Bevorzugung der runden Altersangaben. Dagegen würde eine besonders starke Besetzung der männlichen, dienstpflichtigen Altersklassen die Frage nahelegen, ob der betreffende Ort Garnison hat. In anderen Fällen wird sich als Ursache der auffallenden Form der Kurve ein Rechen- oder Druckfehler oder eine Veränderung der Erhebungsmethode herausstellen, Umstände, die in einer Tabelle leichter verborgen bleiben können. Die graphische Darstellung führt so zu einer auf das Wesentliche beschränkten Kontrolle der Richtigkeit der Zahlen und der Rechnung.

Da es sich im Folgenden um die Methode der Verarbeitung des Zahlenmaterials handelt, sind die Zweifel an der Genauigkeit der Zahlen nicht ausgeführt, außer in den Fällen, in denen sie sich aus den Diagrammen selbst ergeben. Daß die Steuereinschätzungen vielfach hinter der Wirklichkeit zurückbleiben, ist zwar bekannt, nicht aber das Maß der Abweichung. Um daher Schlüsse auf die Verteilung und Veränderung der Einkommen zu ziehen, müßte man wissen, in welchen Klassen der Bevölkerung die Veranlagungen mehr hinter der Wirklichkeit zurückbleiben, und ob und wie stark die Strenge der Einschätzungen sich verändert hat. Mit dieser Einschränkung bleibt die Einschätzung ein wichtiges Symptom der Einkommensverteilung. Durch diese Bedenken wird die Darstellung der veranlagten Einkommen (kurz „Einkommen" schlechthin genannt) nicht behindert, sich der im Anfang (S. 50 ff) erwähnten Methoden zu bedienen.

Betrachten wir nun genauer die Ableitung der erwähnten Pareto'schen Formel für die Verteilung der Einkommen. Diese lautet:

$$N = \frac{H}{R^\alpha}$$

wobei N die Zahl der Personen bedeutet, die das Einkommen von h oder mehr bezeichnen und H und α von Land zu Land wechselnde, für jedes Land aber konstante Größen sind.

Um diese Formel zu finden, zeichnen wir zunächst die Merkmalskurve der Einkommen; wir beginnen mit dem größten Einkommen und tragen das Maximum auf der Ordinatenachse auf, sodann bestimmen wir den Punkt mit der Ordinate für die erste Grenze und der Abszisse für die Zahl der Personen, deren Einkommen diese Grenze überschreitet. Jede Gruppe enthält so ein Trapez, das an das vorhergehende anschließt und ein Polygon bildet. Betrachten wir diese Kurve, so fällt uns auf, daß sie zuerst außerordentlich steil abfällt und dann immer flacher wird.

Um nun die Beziehung der Ordinaten und Abszisse zu bestimmen, suchen

wir nach einer Kurve mit einer einfachen Beziehung von x zu y, die der empirischen Kurve ähnlich sieht. (Siehe Figur 1.)

Eine solche ist die rechtwinklige Hyperbel der reziproken, d. h. umgekehrt porportionalen Werte, deren Gleichung lautet: $y = \frac{1}{x}$; oder allgemeiner $y = \frac{c}{x}$ wobei c eine Konstante ist. Auch diese fällt zuerst außerordentlich steil ab und läuft dann flach aus. Wir können also vermuten, daß für die Einkommenverteilung eine ähnliche Gleichung gilt. Um die Anordnung der Benennungen der Pareto'schen Formel anzupassen, müssen wir die Ordinaten mit x, die Abszissen mit y bezeichnen; die Figur bleibt dabei unverändert, denn wenn

$$y = \frac{c}{x} \text{ ist, ist } xy = c \text{ und } x = \frac{c}{y}$$

Um nun zu untersuchen, ob die wirkliche Einkommensverteilung dieser Formel entspricht, müssen wir aus der Gleichung

$$y = \frac{c}{x}$$

eine solche ableiten, deren Kurve zu einer geraden Linie wird. Die allgemeine Form einer solchen „linearen Gleichung" ist $y = ax + c$.

Um die Gleichung der Hypothese

$$y = \frac{c}{x}$$

in eine linare zu verwandeln, können wir uns der Logarithmen bedienen. Es ist nämlich

$$\log y = \log \frac{c}{x} = \log c - \log x.$$

In dieser Form erhalten wir, wenn wir log y auf der Abszissenachse $\log \frac{c}{x} = \log c \log x$ auf der Ordinatenachse abtragen, eine gerade Linie. (Siehe Fig. 5.)

Auch die Kurve der Einkommen nähert sich bei logarithmischer Behandlung einer geraden Linie, ohne völlig in eine solche überzugehen; noch stärker weicht sie dadurch von der hypothetischen Kurve ab, daß sie nicht im Winkel von 45° geneigt ist.

In letzterer Beziehung kann man die Hypothese der Wirklichkeit dadurch anpassen, daß man alle Abszissen im gleichen Verhältnis verändert, indem man alle log y mit enicm konstanten Faktor α multipliziert. Man erhält dann eine gerade Linie, deren Abszissen $= \alpha \log y$ sind und die mit der Hauptrichtung der empirischen Figur zusammenfällt. Für sie gilt die Gleichung

$$\alpha \log y = \alpha (\log c - \log x)$$

die aus der Gleichung

$$\log y = \log c - \log x$$

durch Multiplikation beider Seiten mit α entsteht.

Nun ist $\alpha \log y$ nach der Annahme $= \log N$; nennt man ferner $c^\alpha = H$, so folgt aus

$$\alpha \log y = \alpha (\log c - \log x)$$
$$\log N = \log c^\alpha - \alpha \log x$$
$$\log N = \log H - \alpha \log R$$
$$\log N = \log \frac{H}{R^\alpha}$$

und ergibt sich durch Delogarithmieren die Pareto'sche Formel

$$N = \frac{H}{R^\alpha}$$

Die Pareto'sche Formel ist also der algebraische Ausdruck der Hypothese, daß die Kurve der Einkommen bei logarithmischer Behandlung der Abszissen und Ordinaten zu einer geraden Linie wird, oder sich einer solchen wenigstens nähert. Die logarithmische Behandlung dient also dazu, die auffallendste Wendung der Kurve zu beseitigen; allerdings ist damit noch nicht jede Krümmung beseitigt.

Betrachtet man nämlich einige solche doppelt logarithmische Kurven der Einkommensverteilung, so fällt auf, daß die größeren Einkommen eine nahezu gerade Linie bilden, die kleineren dagegen weniger zahlreich sind, als die Formel erfordern würde. Dies tritt besonders bei Statistiken hervor, die, wie die sächsische, auch die steuerfreien kleinen Einkommen umfaßt.[1]) Es wäre erwünscht, die Formel so umzugestalten, daß sie auch auf die kleinen Einkommen paßt.

[1]) Siehe Figur.

In dieser einfachsten Gestalt gilt überhaupt die Pareto'sche Gleichung nicht unbegrenzt. Schon ein Blick auf eins der Diagramme genügt, um zu zeigen, daß die der Gleichung entsprechende gradlinige Fortsetzung den Logarithmus der Gesamtbevölkerung an einem Punkt erreicht, der weit höher liegt, als der des geringsten Einkommens eines Individuums; dasselbe gilt für die Zahl der Wirtschaftseinheiten, wo diese die Einkommensteuer zu entrichten haben. Man braucht dabei nicht so weit zu gehen, um darauf hinzuweisen, daß für ein Einkommen = 0 die Zahl der Einkommensempfänger $= \infty$ sein müßte, worauf Pareto erwidern könnte, daß eine Person ohne Einkommen nicht leben kann.

Schon bei einem höheren Einkommen muß die Paretosche Linie sich zur Achse wenden, wovon man sich überzeugen kann, wenn man den Logarithmus der Gesamtzahl der Bevölkerung auf der Figur einträgt. Über die Gestaltung dieses unteren Teils der Einkommenskurve gibt die Steuerstatistik nur sehr unvollkommene Kunde. Eine Ausnahme macht Sachsen, dessen Statistik über alle Personen berichtet, die überhaupt ein selbständiges Einkommen aus Arbeit oder Besitz beziehen. Zeichnen wir hiernach ein Pareto'sches Diagramm, so zeigt sich deutlich, wie die Linie sich umbiegt. So z. B. in Oldenburg, Sachsen-Weimar, Norwegen (s. Fig. 4), in italienischen Gemeinden.

Auch in manchen anderen Ländern zeigt die Kurve eine geringe Krümmung dadurch, daß die kleineren Einkommen weniger zahlreich sind, als der Gleichung $\log N = \log H - \alpha \log R$ entsprechen würde.

Um die Formel dieser Tatsache besser anzupassen, bedient sich Pareto einer dritten Konstante a und schreibt:

$$N = \frac{H}{(R + a)^{\alpha}}; \log N = \log H - \alpha \log (R + a)$$

Er erreicht damit, daß die Abweichung $\Delta \log N$ der auf Grund des Mittelwerts für α berechneten und der empirisch gefundenen Werte für log N geringer ist, als bei der einfachen Formel.

Diesem Vorteil stehen jedoch erhebliche Nachteile gegenüber. Während für die einfache Formel die Konstante α einen unmittelbar vergleichbaren Wert hat, und in einer einfachen Beziehung zur Konzentration der Einkommen steht, läßt sich bei der Gleichung mit 3 Konstanten eine solche Beziehung nicht unmittelbar erkennen. Vielmehr ist die Wirkung von a der jenigen von α entgegengesetzt, wie sich aus der Figur ergibt. Eine Vergleichung wird nur möglich, indem man einen Mittelwert für α unter Eliminierung von a berechnet, also auf die Gleichung mit 2 Konstanten zurückgeht.

Dazu kommt, daß die Berechnung der Werte von H, α und a sehr kompliziert wird, ebenso die Zeichnung der Figur und die Vergleichung der empirischen Daten mit der interpolierten Kurve und Zahlenreihe. Diese Nachteile könnten noch in den Kauf genommen werden, wenn durch Einführung von a ein positives Ergebnis erreicht würde. Das ist aber nicht der Fall. Denn da die Krümmung der Kurve, die durch a berücksichtigt werden soll, vor allem in den unteren Einkommensklassen eintritt und ebenfalls durch a nur annäherungsweise wiedergegeben wird,[2]) ist die Größe von a davon abhängig, bis zu welchem

[2]) Deren Methode bei Pareto meines Wissens nicht angegeben ist.

Minimaleinkommen sich die Einkommensstatistik erstreckt. Ohne dessen Angaben liefert die Berechnung von a[3]) keine vergleichbaren Werte für eine wirt-

[3]) Bei den mittleren Einkommen, an der Grenze der Arbeits- und Besitzeinkommen, tritt häufig sogar eine konvexe Stelle, ein Knick, ein.

schaftlich erhebliche Tatsache, sondern zeigt nur, wie weit für die betreffende Statistik sich die erste Pareto'sche Gleichung von der zweiten unterscheidet.

Dasselbe gilt von der vierten Konstante b in der Formel

$$\log N = \log H - \alpha \log (R + a) - b R$$

durch welche eine Abweichung der Häufigkeit der größeren Einkommen berücksichtigt werden soll. Auch ist sie nur ein Hindernis auf die Ungenauigkeit der Formel.

Diese Bedenken werden beseitigt, wenn man darauf verzichtet, die empirische Mannigfaltigkeit durch einfache analytische Gebilde voll erfassen zu wollen. Betrachtet man die Kurve, wie sie wirklich ist, so zeigt, sich daß α das Verhältnis der Differenzen der log N zu denen der log R, seinen Wert beständig, wenn auch in der Regel nur geringfügig ändert. Man kommt daher der Wirklichkeit am nächsten, wenn man α nicht als Konstante, sondern als Maß dieses Verhältnisses[1]) auffaßt und die Werte von α angibt, die sich un-

[1]) Trigonometrisch ist $\alpha = \text{tang}$ (log N, log R).

mittelbar aus der Berechnung oder aus der Figur ergeben, und zwar zunächst für jede Einkommensklasse von einer Stufe der Statistik bis zur nächsten. Will man die Resultate vereinfachen und vergleichbar machen, so dürfte es sich empfehlen, für gleichmäßige Prozentsätze der Bevölkerung den Wert von α zu berechnen. Z. B. für die perzentilen Grade: 0,1, 0,5, 1 ,5, 10 und ev. 50% oder für 0,1, 0,2, 0,5, 1, 2, 5, 10, 20 und ev. 50%.

Auf den logarithmischen Diagrammen läßt sich diese Einteilung leicht anbringen. Genauer würde es sein, die erwerbstätige Bevölkerung zu Grunde zu legen, dies ist aber nur soweit möglich, als für diese vergleichbare Daten vorliegen.

Aus diesen einzelnen Werten für α ließen sich einerseits vergleichbare Mittelwerte für verschiedene Länder berechnen, anderseits konnte man genauer ersehen, wie sich α innerhalb derselben Bevölkerung ändert. Für praktische Zwecke wird es sich aber in beiden Fällen empfehlen, wenn möglich anstatt α einen andern Wert zu suchen.

Handelt es sich darum, mit einem einzigen Ausdruck das Maß der Konzentration zu kennzeichnen, so eignet dazu besser Gini's δ; soll dagegen der Unterschied der einzelnen Einkommensklassen festgestellt werden, so eignet sich dafür besser der Wert y in der ersten Gleichung Pareto's[1]) oder Y, die nicht

[1]) La legge della domanda, 1905.

kumulative Häufigkeit für Intervalle des natürlichen Logarithmen.

Was zunächst den Gini'schen Index der Konzentration betrifft, so führt

Gini diesen Begriff anstelle dcs von Pareto xeigenartigen definierten Begriffs der inégalité ein, der zu Mißverständnissen Anlaß gegeben hatte.[2])

[2]) Pareto hatte in § 964 zunächst ausgeführt, daè man wohl zu unterscheiden habe zwischen einer Verminderung des Pauperismus und einer Verminderung der Ungleichheit. Nur die erstere bedeute eine wirkliche Verbesserung der Lage der Masse der Bevölkerung; daher sei nur eine solche Verminderung der Ungleichheit als solche zu bezeichnen, die auf eine Hebung derd kleineren Einkommen und daher auf einer Verminderung der Gliederungszahlen der ärmeren Bevölkerung unter den verschiedenen Einkommnsgrenze beruhe. Im folgenden identifiziert Pareto seine Difinition der Verminderung der Ungleichheit mit dieser Bedingung und schreibt (§ 964, Schluß), daß für die folgenden Ausführungen als diminution de l' inégalité gelten solle. wenn $\frac{U x}{U n}$ (d. i. die Zahl der Personen und Einkommen über x als Bruchteil alle Censiten über der unteren Grenze n) wachse also der Anteil der Ärmeren $\frac{Un - Ux}{U n}$ abnehme. Diese künstliche Definition, die die relative Verminderung des Pauperismus und der ärmeren Klassen als Verminderung der Ungleichheit bezeichnet. führt zu Mißverständnissen, zumal wenn sie in Aufsätzen ohne nähere Erläuterung angewendet wird. Dazu kommt, daß unter der Voraussetzung der Geltung der Gleichungen

$N = \frac{H}{R^{\alpha}}$ oder $N = \frac{H}{(R + a)^{\alpha}}$ eine „diminution de l'inégalité eintritt, wenn α abnimmt oder a zunimmt, während eine Verminderung der Ungleichheit, die nach Bortkiewicz (Schmollers J.B. 1898) eher der Verminderung der durchschnittlichen einfachen oder quadratischen Abweichungen vom mittleren Einkommen entsprechen würde, gerade im umgekehrten Fall eintritt. Will man also die Ungleichheit durch α messen so deutet nach dem allgemeinen Sprachgebrauch eine Zunahme von α auf eine Verminderung der Ungleichheit. Bresciani hält im Gegensatz zu zu Pareto 1 für einen „quantitativen summarischen Ausdruck der Ungleichmäßigkeit der Einkommensverteilung", indem diese um so gleichmäßiger ist, je größer $\frac{1}{a}$ ist. Vergl. Vogel a. a. O. S.511. Tabacovi verwendet als Index der Ungleichheit den Quotienten des Durchschnittseinkommens des reichsten Prozents der Ceniten durch das Durchschnittseinkommen der übrigen 99 %.

Dieser Quotient ist $\frac{99}{100^{\frac{1}{\delta}} - 1}$ also nahezu $= 100^{1-\frac{1}{\delta}} = 100^{\frac{1}{\delta}}$

Gini geht davon aus, daß die Konzentration der Einkommen zunimmt, wenn sich ein größerer Bruchteil der Einkommen in den Händen eines geringeren Bruchteils der Bevölkerung bezw. der Censiten befindet. Diese Defination dürfte wohl kaum Widerspruch finden. Als kurzen Ausdruck der Konzentration in diesem Sinne, als „indice die concentrazione", wählt Gini den Exponenten mit dem man den Quotienten des reicheren Teils durch die Gesamtzahl der Censiten potenzieren muß, um den Quotienten ihres Einkommens durch das Gesamteinkommen zu erhalten. Nennt man die Gesamtzahl der Censiten n, irgend eine Zahl der Reicheren m, die Summe ihrer Einkommen

$$\sum_{i=n}^{i=1} a_i \quad \text{bzw.} \quad \sum_{i=n}^{i=n-m+1} a_i$$

so ist

$$\left(\frac{\sum_{i=n}^{i=n-m+1} a_i}{\sum_{i=n}^{i=1} a_i} \right)^{\delta} = \frac{m}{n}, \text{ und } \delta \text{ der}$$

Index der Konzentration. Dabei bedeutet i die Ordnungszahl eines Censiten, beginnend mit dem ärmsten und a, das Einkommen des betreffenden Censiten.

Da für alle Einkommengrenzen eines Landes die Summe der Einkommen und Censiten aller höheren Klassen mit den Gesamtzahlen aller Censiten und

Einkommen verglichen werden, bleiben letztere konstant. Man erhält also für m die Gleichung

$$m = \left(\sum_{i=n}^{i-n-m+1} a_i \right)^{\delta} \cdot \frac{1}{K}$$

oder logarithmisch

$$\log m = \delta \log \left(\sum_{i=n}^{i=n-m+1} a_i \right) - \log K$$

Nach der hier angewandten Bezeichnungsweise lautet diese Gleichung

$$\log N = \delta \log A - \log K.$$

Gini unterscheidet zwischen „indici semplici" mit einer und „indici complessi" mit mehreren charakteristischen Konstanten, ferner „indici medi", die nur Mittelwerte für einen ganzen Kollektivgegenstand sind, und „indici descrittivi" oder „coefficienti", die auch für die einzelnen Werte einer Reihe nahezu zutreffen. δ ist demnach ein indice semplice, und für die Einkommensverteilung in der Regel ein indici descrittivo, für die Einkommen aus Besitz dagegen nur ein indice medio.

In der Tat nähert sich die Kurve der log N, log A noch mehr einer geraden Linie las die Pareto'sche der log N, log R und sind dle Abweichungen Δ log N der empirischen und der interpolierten Werte für log N verhältnismäßig geringer.

Dies ist kein Zufall, denn es besteht ein innerer Zusammenhang zwischen beiden Kurven und ihren Gleichungen, und der Aufbau der Gini'schen Linie ist derart, daß die Krümmungen noch mehr ausgeglichen werden müssen, als bei derjenigen Pareto's. (Siehe Fig. 9.)

Um dies nachzuweisen, kann man die kumulativen Kurven Pareto's in einfache Häufigkeitskurven verwandeln und eine entsprechende Linie aus den Gini'schen Daten ableiten. Zieht man in der Pareto'schen Kurve Parallelen zu log A Achse, so entstehen gleiche Intervalle der log R und daher relativ gleiche Intervalle der Einkommen. Die Zahlen der auf diese entfallenden Censiten werden mit D N bezeichnet. Die Größe der D N ist gleich der Differenz der beiden sie begrenzenden N. Den Logarithmus dieser Differenz log D N hat man an der Stelle des durchschnittlichen Einkommens einzutragen.

Um aus den D N die Häufigkeit zu erhalten, muß man sie durch das Intervall D log R der log R dividieren; man hat dieses gleich 0,1 gewählt, so sind sie zu verzehnfachen, allgemein ist $Y = \frac{D N}{D \log R}$ Diese Häufigkeit könnte man in der Größe der natürlichen Zahlen darstellen, wenn man über großes Format verfügt oder darauf verzichten will, die Häufigkeit der größeren Einkommen erkennbar zu machen. Andernfalls ist der Logarithmus der Häufigkeit log Y einzutragen, und zwar an der Stelle des durchschnittlichen Einkommens jeder Klasse. Als solches wäre es falsch, das ungewogene arithmetische Mittel der Einkommensgrenzen anzunehmen, da die kleineren Einkommen zahlreicher sind, als die größeren. Der gewogene Durchschnitt ist kleiner und wird dem geometrischen Mittel der Einkommensgrenzen näher kommen. Dieses ist zugleich leicht zu zeichnen, indem man das Intervall der log R halbiert. Je kleiner man die Intervalle macht, desto größer wird die Genauigkeit. Für unendlich kleine Differenzen der natürlichen Logarithmen, d l R, ist die Häufigkeit $\frac{d N}{d l R} = - \alpha N$. Sie ist negativ, weil mit wachsendem R die N abnehmen. Es sei daher $Y = \frac{- d N}{d l R} = \alpha N$

Ist eine gerade Pareto'sche Linie der genaue Ausdruck der Einkommensverteilung, so ist auch die Häufigkeit für die einzelnen logarithmischen Intervalle durch eine grade Linie darzustellen.

Diese ist (log Y, log R) = (log (α N), log R) (siehe Anhang).

Daraus ergibt sich, daß die für die Einheit der natürlichen Logarithmen zuwachsenden Einkommensbeträge durch eine grade Linie nach der Formel dargestellt werden

$$Z = R\left(\frac{-dN}{d\,l\,R}\right) = R\,Y = \alpha\,.\,N\,.\,R.$$

Die Summen dieser Einkommen sind dann

$$A = \delta\left(\frac{-dN}{d\,l\,R}\right).\,R\,.\,dlR\,. = \frac{\alpha}{\alpha-1}\,.\,R\,.\,N \qquad \text{(siehe Anhang)}$$

Ändert die Pareto'sche Kurve an einem Punkte ihre Richtung, so daß z. B. α für die größeren Einkommen = 2 für die kleineren = 1,5 ist, so werden für den betreffenden Wert von R sowohl A als N kleiner, als sie sein würden, wenn auch für die höheren Einkommen α = 1,5 gelten würde, nämlich

$$N = Y\,.\,\frac{1}{2} \text{ anstatt } N = Y\,.\,\frac{1}{1{,}5}$$

und

$$A = Y\,.\,\frac{R}{2-1} \text{ anstatt } A = X\,.\,\frac{R}{1.5-1}$$

Um diesen Betrag werden auch alle folgenden N und R kleiner, als sie sein würden, wenn α = 1,5 für die ganze Figur gelten würde. Bezeichnet man diese Werte von A, N und R mit Ax, Nx und Rx und α = 2 mit αx, so gilt für die R < Rx

$$N = \frac{Y}{\alpha} - \frac{Yx}{\alpha} + \frac{Yx}{\alpha x} = \frac{Y}{1{,}5} - \frac{Y\,x}{6}$$

$$A = \frac{R\,.\,x}{\alpha-1} - \frac{Rx\;Yx}{\alpha-1} + \frac{Rx\;Yx}{\alpha x-1} = 2\,RY - RxYx$$

Von den folgenden $\frac{X}{\alpha}$ und $\frac{R\,Y}{\alpha 1}$ werden also konstante Größen abgezogen, von denen die letztere bedeutend größer ist. Die Logarithmen werden dadurch vermindert, zunächst so stark, daß sich die Kurve der log A und log N für die kleineren log R an die für die größeren anschließt, dann in abnehmendem Maße. Diese Kurven erhalten so eine erhebliche Krümmung, besonders die der (log A, log R). Trotzdem weicht die (log A, log N) Kurve Gini's nur wenig von einer graden Linie ab. Daraus ergibt sich, daß δ zwar ein vorzüglicher Index der Konzentration ist, aber auch, daß die Gini'sche Kurve ungeeignet ist, die feineren Abweichungen erkennen zu lassen. Dazu kommt, daß eine Verminderung von α für die kleineren Einkommen, also ein geringeres Fortschreiten von log N, bewirkt, daß die entsprechenden Punkte der Gini'schen Kurve nahe aneinanderrücken, so daß die Einkommensentwicklung der unteren Klassen undeutlich wird.

Verfolgt man nicht den Zweck, eine möglichst grade, sondern eine deutlich gekrümmte Linie zu erhalten, so ist die (log Y, log R) Kurve oder die (log Z, log R) Kurve vorzuziehen.[1]

[1] Siehe Anhang!

Erstere ist durch Hinzufügen von log α, aus der log N Kurve Pareto's, letztere durch Addition von log α + log R zu gewinnen.

Bei der Konstruktion entsteht die Schwierigkeit, daß nur für bestimmte Werte von R Angaben über N vorliegen, die aber gerade nur für die dazwischenliegenden Intervalle berechnet werden können. Addiert man diese log α zu dem Durchschnitt der begrenzenden log N, so oszilliert die Kurve stärker, addiert man umgekehrt den Durchschnitt der log α zu dem log N, so oszilliert sie schwächer. Am genauesten scheint mir zu sein, beide Reihen zu kombinieren, wie es auf der Figur für Norwegen durchgeführt ist. Ein auffallendes Bild zeigt das Diagramm von Triest 1904. (S. Fig. 5.) Die großen Einkommen haben große

Intervalle und zeigen das verschiedene Ergebnis der Berechnungsarten von Y. Man sieht, wie schwache Abweichungen der Pareto'schen Kurven zu starken Ausbuchtungen werden, und daß die Methode log Y = log $\alpha + \frac{\log No + \log N1}{2}$ zu setzen, die Anschwellungen der log Y Kurve besser mit den flachen Stellen der log N Kurve korrespondieren läßt. Die Zusammenfassung je zweier ganzen Intervalle zur Berechnung von α führt bei ihrer großen Ungleichheit zu keinem günstigen Ergebnis. Für die kleineren Einkommen sind die Intervalle sehr klein. Die Folge ist ein lebhaftes Oszillieren der log Y Kurve. Einige Maxima treten hervor, ein besonders starkes in der untersten Steuerstufe. Dies scheint die Bemerkung Gini's[1]) zu bestätigen, daß die Steuerbehörden eher geneigt sind, zu einer geringeren Steuer einzuschätzen, als ganz auf die Steuer zu verzichten. Es wäre sehr interessant, an der Hand einer Lohnstatistik darüber Anhaltspunkte zu gewinnen, wie die Kurve unterhalb der Steuergrenze weiterverläuft und es ist bedauerlich, daß das Material, das aus den Angaben der Arbeitgeber über die Löhne gewonnen werden könnte, nicht in die Einkommensstatistik aufgenommen wird.

[1]) Zitiert bei Furlan, Neuere Literatur zur Einkommenverteilung.

Für diese, anscheinend auch bei logarithmischer Skala der Einkommen am dichtest besetzten Einkommensklassen sind wir auf Lohnstatistiken angewiesen. Eine Ausnahme machen Sachsen und Bayern.

Die Häufigkeitskurven der Arbeitslöhne hat Lucien March durch die Formel interpoliert[1])

[1]) Siehe S. 59.

$$y = \frac{A\,R^{m}}{{}_{a}R}$$

Zwei Kurven nach dieser Gleichung bringt nebenstehende Figur 8. Sie sind zwar stark asymmetrisch, brechen aber doch nicht beim Minimum plötzlich ab, wie es Pareto für die Häufigkeitskurve der Einkommen vermutete.[2])

[2]) Cours d' éc. pol. II § 961, Abs. 4.

Sie wurden um so breiter, je größer m ist; log A verschiebt die Abzissenachse nach unten, während a Einfluß auf die Form der Kurve hat. Je größer log a ist, desto steiler fällt die Kurve auf der rechten Seite, für die größeren Werte von R ab.

Die hier angewandte einfache Konstruktion hat die Eigentümlichkeit, daß die Skala der Häufigkeit logarithmisch, die der Merkmale aber arithmetisch ist. Will man die Figur beiderseitig logarithmisch machen, so muß man anstatt der Kurve

$$(\log y,\ R)$$

diejenige

$$(\log Y,\ \log R)$$

zeichnen, wobei

$$Y = \frac{-\,d\,N}{d\,l\,R} \text{ ist.}$$

Aus den Differenziationen (siehe Anhang) ergibt sich, daß log Y = log y + log R ist.

Man kann jeweils log R addieren, oder für die Formel

$$\log Y = \log y + \log R = \log A + m \log R \times R \log a + \log R$$
$$= \log A + (m+1) \log R \times R \log a$$

eine Figur zeichnen. Man sieht, wie die logarithmische Umwandlung der R-Achse die linke Seite der Kurve auseinanderzieht und erniedrigt, die rechte zusammendrängt und erhöht. Das Maximum liegt so bei den größeren log R, und die Gestalt der Kurve wird für log a = 1 und = 2 sehr ähnlich.

Die Formel Lucien March's gestattet es zwar, durch einen kurzen Ausdruck eine Reihe von Lohndaten zusammenzufassen und ihre Besonderheiten zu messen; sie gibt aber eine klare Anschauung nur dann, wenn man sich die kombinierte Wirkung der drei Konstanten durch eine Zeichnung oder durch eine besonders starke Vorstellungkraft vergegenwärtigt. Eine solche Zeichnung kann nur die bereits besprochenen Formen annehmen.

Gehen wir nun von den Darstellungsformen zu den erklärenden Hypothesen. L. v. Bortkiewicz sagt hierüber in seiner Besprechung des Pareto'schen Werkes:[1])

[1]) Schmollers Jahrbuch 1898. IV. S. 118.

„Ich will hier von der Frage absehen, ob und in welchem Grade die gefundenen Ergebnisse der Rechnung danach angetan seien, die Behauptung Pareto's zu bestätigen, daß die Einnahmekurve überall gleiche Gestalt aufweist. Gesetzt, daß letzteres der Fall sei, was wäre mit einer solchen Erkenntnis gewonnen? Recht viel, wenn man imstande wäre, zu zeigen, daß die empirisch gefundene Formel zugleich aus gewissen apriorischen Erwägungen herzuleiten ist. Äußerst wenig, wenn letzteres nicht zutrifft. Im ersteren Fall ließe nämlich die Übereinstimmung mit einer gewissen Wahrscheinlichkeit darauf schließen, daß jenen apriorischen Erwägungen bezw. Voraussetzungen etwas reelles entspricht, obwohl die Schlußfolgerung keine zwingende wäre, weil sich ja dasselbe Resultat aus verschiedenen Prämissen ergeben kann. Je größer aber die innere Wahrscheinlichkeit der erwähnten Voraussetzungen sein würde, um so schwieriger wäre es für den Verstand, sich jener Schlußfolgerung zu entziehen. Auf diese Weise vermag manchmal die Tatsache, daß sich gewisse statistische Daten in eine mathematische Formel fügen, uns über die Wirklichkeit zu belehren. Wendet man nun einen ähnlichen Maßstab auf Pareto's Gesetz an, so kommt man zu der Überzeugung, daß dasselbe einer tieferen Bedeutung entbehrt. Wohl sucht Pareto von gewissen mehr oder weniger annehmbaren Voraussetzungen heraus seine Einkommenkurve begreiflicher zu machen. Diese falle mit dem mathematischen Fehlergesetz nicht zusammen, bemerkt Pareto. Also sei die Verteilung der Einkommen nach ihrer Größe nicht für ein Ergebnis des reinen Zufalls zu halten. Pareto neigt vielmehr zu der Auffassung, daß sich in den Größeverschiedenheiten der Einkommen die Ungleichheit der Veranlagung der Menschen abspiegelt. Konstruiert man eine Kurve, deren Ordinaten angeben, wie sich ein Tausend von Menschen bezw. von Geborenen nach dem Maß der Fähigkeit verteilen, ein Einkommen von bestimmter Höhe zu erzielen, oder, genauer ausgedrückt, nach der Größe des Einkommens, welches zu erzielen jene Menschen kraft ihrer Eigenschaften befähigt sind, so sei es, nach Pareto, sehr wahrscheinlich, daß die Gestalt solch einer Kurve von Land zu Land und von Epoche zu Epoche nicht wechselt, weil hierbei die Natur des Menschengeschlechts ausschlaggebend sei (§§ 1012, 1026). Auf diese Weise wird eine nicht ohne weiteres verständliche Erscheinung (das Gesetz der Einkommenverteilung) auf zwei Hypothesen zurückgeführt, die eine darin bestehend, daß es ein bestimmtes Gesetz der Verteilung der Menschen nach ihrer Fähigkeit gibt, Einkommen von bestimmter Höhe zu erzielen, die andere in

dem Satz gipfelnd, daß die tatsächliche Verteilung der Einkommen nach ihrer Größe durch das Maß jener Fähigkeit ausschließlich bestimmt werde.[1]) Im

[1]) Das Folgende steht in einer Anmerkung.

Anhang auf S. 416—419 sucht Pareto die Verteilung der Einkommen nach ihrer Größe als ein Ergebnis der kombinierten Wirkung von zwei Faktoren hinzustellen: nämlich 1. der soeben erwähnten verschiedenen Befähigung der Menschen und 2. des „Zufalls". Jedoch werden bei dieser Ausführung für die Wirkung der „zufälligen Ursachen" Bedingungen gestellt, welche im Endresultat zu einer gänzlichen Eliminierung des Zufalls führen, und so erscheint die Heranziehung der Wahrscheinlichkeitsrechnung als ziemlich überflüssig. Es läuft einfach darauf hinaus, eine Verteilung der Menschen nach dem Maß der bewußten Fähigkeit vorauszusetzen, welche dem gefundenen Gesetz der Verteilung der Einkommen genau entspricht."

In der Tat hat es nicht an Hypothesen gefehlt, die die Regelmäßigkeit der Pareto'schen und der ihr ähnlichen Kurven kausal zu erklären versuchten.

Pareto selbst gibt im Anhang die von Bortkiewicz erwähnte Hypothese. Er nimmt an, daß sich die Menschheit in Gruppen teilen lasse, für die Wahrscheinlichkeit eine arithmetische Reihe von unendlich kleinen Abstufungen zwischen 0, d. h. der Unmöglichkeit und 1, d. h. der Gewißheit durchlaufen. Im Durchschnitt einer längeren Periode verteilen sich dann die Einkommen innerhalb jeder Gruppe nach dem Gesetz der großen Zahl, so daß die größte Häufigkeit für ein Einkommen gemäß der betreffenden Wahrscheinlichkeit eintritt und die Häufigkeit im übrigen asymmetrisch nach beiden Seiten abnimmt; für jedes Einkommen besteht also innerhalb jeder Gruppe eine bestimmte relative Häufigkeit, deren absoluter Wert von der Größe der Gruppe selbst abhängt. Pareto schreibt dann: „es handelt sich jetzt darum, für g (d. h. die Größe der Gruppen) eine solche Funktion zu finden, daß man für genügend große Werte von x (den Einkommensgrenzen) das Gesetz erhält, das wir experimentell gefunden haben." Dies Gesetz wird ausgedrückt durch die Gleichung

$$y = \frac{A}{x^{h}}; \text{ d. h.: } y = \frac{H^{\alpha}}{R^{\alpha+1}} \quad ^{1})$$

[1]) Links Bezeichnungsweise Paretos, rechts Bezeichnungsweise dieser Arbeit.

Im Folgenden berechnet Pareto für die g eine Funktion, die mit derjenigen für y zusammenfällt.

Pareto hat somit nachgewiesen, daß die Größe der Gruppen ebenso abgestuft ist, wie die der Einkommensklassen, wenn die Einkommensbildung in der hypothetischen Weise abgestuft ist. Man würde also die Erwerbsfähigkeit des Menschen an dem Erfolg messen können, unter der Voraussetzung, daß der Erfolg sich aus einer Summe gleicher Einheiten zusammensetzt, die mit verschiedener Wahrscheinlichkeit gewonnen werden. Durch diese Hypothese ist aber aus zwei Gründen keine kausale Erklärung gegeben. Erstens ist damit die beobachtete Tatsache nicht auf eine bekannte oder wenigstens plausible Erscheinung zurückgeführt, zweitens, es ist ebensowenig an sich einleuchtend, warum die Begabungen der Menschen sich nach der Pareto'schen Formel gruppieren, als ihre Einkommen, noch ist es meßbar, ob sie es tun, da man ja nicht weiß, ob der Einzelne einen hohen Platz in einer niedrigen Gruppe oder einen niedrigen Platz in einer hohen Gruppe einnimmt, d. h. ob ihn das Glück begünstigt hat oder die Begabung.

Ferner scheint mir auch die verbindende Hypothese mit der Wirklichkeit nicht im Einklang zu stehen. Die Menschen zerfallen nicht nach ihren wirt-

schaftlichen Fähigkeiten in Gruppen, innerhalb deren jeder Einzelne die Möglichkeit beständigen Erfolgs und beständigen Mißerfolgs hätte, wobei sie nur die verschiedene Mischung der Erfolge und Mißerfolge innerhalb der Gruppen um Zentren der Häufigkeit zusammendrängen, die je nach der Gruppe verschieden gelagert sind. Noch weniger sind diejenigen Gruppen am zahlreichsten, bei denen die Wahrscheinlichkeit des Erfolgs am geringsten ist.

Vielmehr ist die Gruppierung der Menschen derart, daß der einzelne wirtschaftliche Erfolg (nicht im Ganzen, sondern von Fall zu Fall), also ein Erfolg, der etwa dem Erwerb von 1 Fr. entsprechen würde, in der Mehrzahl der Fälle tatsächlich eintritt und seine Größe nicht von der Mischung von Erfolg und Mißerfolg, sondern von dem Besitz der Vorbildung und der sozialen Stellung abhängt. Die Mehrzahl der Menschen lebt in der Regel in gesicherter Existenz oft auf Generationen, auf Jahre oder zum mindesten auf Wochen hiaus; nur wenige sind es, bei denen jeder einzelne Erwerb vom Zufall abhängt und eine Regelmäßigkeit nur für den Durchschnitt einer großen Anzahl von Fällen eintreten kann. Eine solche Unsicherheit des Erfolges trifft man nicht nur, wie es der Pareto'schen Hypothese entsprechen würde, bei den niedrigsten Einkommen, den Gelegenheitsarbeitern, Bettlern, Verbrechern, sondern gerade auch bei den eventuell sehr hohen Einnahmen von Spekulanten, Künstlern, Erfindern.

Immerhin hat Pareto mit dieser Hypothese den Standpunkt verlassen,[1]

[1]) Vergl. § 962 und Annexe zu § 962.

daß die Regelmäßigkeit seiner Diagramme den „Zufall" und damit das Gesetz der Wahrscheinlichkeit als erklärende Ursache ausschließe.

In der Tat scheint mir das Umgekehrte richtig; würden die Diagramme verschiedener Länder und Zeiten in der Form stark abweichen, so könnte man auf bestimmte Ursachen schließen. Diese könnte man zwar vom Standpunkt der sonst beobachteten Regelmäßigkeit als „Zufälligkeiten" bezeichnen; nicht aber im Sinne des Gesetzes der Wahrscheinlichkeit. Denn bei diesen handelt es sich nicht um große, deutlich hervortretende Einwirkungen, sondern um die kleinen abändernden Ursachen, die wegen ihrer Massenhaftigkeit im einzelnen unverkennbar sind. Das Gauß'sche Gesetz und die „courbe des probabilités" zeigen, wie diese sich bei ihrem Zusammenwirken in praktisch unbegrenzter Zahl gegenseitig zum Teil kompensieren, und so wieder eine Regelmäßigkeit entstehen lassen.

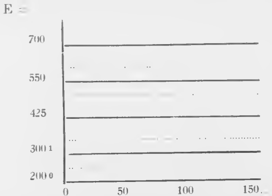
Solche unendlich viele und kleine abändernde Ursachen wirken auf das menschliche Erwerbsleben ein, aber nicht in der Weise, daß es von ihnen abhinge, ob jemand bei jeder einzelnen Gelegenheit 1 Fr. erwirbt oder nicht erwirbt, sondern ob jemand seine wirtschaftliche Lage verbessert oder verschlechtert und ob die vorbereitenden Bedingungen zu einem solchen entscheidenden Schritt geschaffen oder wieder beseitigt werden.

Diesem Gedanken trägt V. Furlan[1]) in der Weise Rechnung, daß er aus

[1]) V. Furlan, Note sulla nuova Paretiana dei redditi. Giornale degli Economisti.

der Konkurrenz der Bestrebungen aller nach Aufsteigen auf der wirtschaftlichen Stufenleiter, einen Gleichgewichtszustand entstehen läßt, der sich mit der Verteilung nach der Pareto'schen Formel deckt.

Furlan ordnet die Censiten nach Einkommensklassen und innerhalb dieser nach der Dauer der unnuterbrochenen Zugehörigkeit zu einer Einkommensklasse etwa in folgender Form (siehe Figur), wobei ξ das mittlere Einkommen der



Klasse der Censiten und u seine Nummer unter den nach der Dauer des ununterbrochenen Besitzes eines Einkommens von den betreffenden E oder mehr bedeutet. Der Rang jedes Individuums ist demnach durch E und u bestimmt.

Auf die Veränderung dieser Rangordnung wirkt die Gesamtheit der menschlichen Eigenschaften ein, von denen Furlan nur die vererbten, nicht die erworbenen, betrachten will, und zwar unter der Voraussetzung vollkommen freier Konkurenz.

Furlan gibt dann einige Hypothesen und beweist, daß bei deren Zutreffen eine der Pareto'schen Formel $\log N = \log H - \log (R + a)^{\alpha}$ entsprechende Verteilung der Einkommen resultieren würde.

Leider hat er diese Hypothesen nur in der Weise ausgedrückt, daß er die Kraft sf jedes Individuums, die Konkurrenz der anderen Individuen der Gesamtheit zu überwinden durch komplizierte Gleichungen mißt. Dadurch bleibt der reale Inhalt seiner Hypothesen für den Nichtmathematiker unverständlich. Furlan läßt sf von dem Unterschied des Ranges und gewissen Konstanten abhängig sein, und nimmt diese Konstante für alle Individuen der Gesamtheit als gleich an. Es scheint mir aber, daß man weder Elemente, die auf die Kraft der Individuen von Einfluß sind, sich im Kampf um eine bessere materielle Existenz zu behaupten, für alle Individuen einer Gesamtheit als gleich annehmen darf, noch daß sich die Konkurrenz der Individuen außer auf die Erreichung einer höheren Einkommensstufe auch darauf bezieht, innerhalb dieser Stufe einen höheren Rang der Anciennität zu erlangen. Durch diese Hypothesen wird auch nur die allgemeine Formel Pareto's $N = \frac{H}{(R+a)^{\alpha}}$ erklärt, die, wie wir gesehen haben, je nach der Größe von a einer Kurve von sehr verschiedener Krümmung entsprechen kann.

Es liegt nahe, die Einkommensverteilung mit der Anordnung der Häufigkeit zu vergleichen, die sonst beobachtet worden ist. Fechner[1]) ist in seiner Kol-

[1]) Fechner, Kollektivmaßlehre S. 333.

lektivmaßlehre zu dem Ergebnis gekomen, daß in der Regel das zweiseitige Gauß'sche Gesetz gilt, d. h. daß sich die Häufigkeiten verhalten, wie die Grenzwerte der Koeffizienten einer binomischen Entwicklung, die von dichtestem Wert ab unsymetrisch sich über die Grade des Merkmals erstrecken. Da der dichteste Wert von der Einkommensstatistik in der Regel nicht erfaßt und nie erheblich überschritten wird, kann von der Asymetrie und den feineren theoretischen Ausgestaltungen hier abgesehen werden. Wichtig ist aber die weitere Beobachtung Fechners, daß in der Regel die relativen, nicht die arithmische Skala der Merkmalsgrade anzuwenden ist, und zwar besonders wenn

soluten Unterschiede des Merkmals von Bedeutung sind, daß also eine logarith-
diese Grade erheblich von einander abweichen. Dies trifft im höchsten Maße bei der Einkommensverteilung zu. Die höchsten veranlagten Einkommen sind mehr als 1000fach größer als die niedrigsten, und unter dieser setzt sich die Reihe der Abstufungen bis zum äußersten Existenzminimum fort.

Will man daher die „courbe des revenus" mit der „courbe des probabilités" vergleichen, so darf man nicht für letztere eine arithmetische Skala wählen, wie es Ammon[2]) und auch Pareto[3]) tun. Allerdings behält Pareto auch für

[2]) Ammon, Gesellschaftsordnung S.

[3]) Pareto, Cours d' économie politique, II § 962.

eine logarithmische Skala der Einkommen mit seinem Ergebnis recht, daß seine Gleichung und Kurve von dem Gesetz der zufälligen Abweichungen verschieden ist.

Um beide zu vergleichen, ist es am leichtesten, die sogen. Fehlerkurve ebenfalls kumulativ und beiderseits logarithmisch zu gestalten. (S. Fig. 7.) Sie ist dann ein nahezu symmetrisches Gegenstück zur Logarithmuslinie, indem sie beim Maximum ins Negativunendliche abfällt, wie die Logarithmuslinie bei $\log 0 = -\infty$. Ihr oberer Teil zeigt dagegen eine gewisse Ähnlichkeit mit manchen Pareto'schen Diagrammen, indem sie eine leichte, nach oben hin abnehmende Krümmung zeigt. Man könnte versucht sein, die Unterschiede durch zufällige Einwirkungen oder Beobachtungsfehler zu erklären.

Zeichnet man jedoch nicht kumulative Diagramme der Einkommensverteilung, so treten die Unterschiede deutlicher hervor. Während die logarithmische, nicht kumulative Fehlerkurve einer Parabel nahekommt,[1]) ist die Kurve

[1]) Ihre approximative Gleichung $y = A - e\, h^2 x^2$ ist die einer Parabel.

der Einkommen in der Regel zwar nicht mehr eine Grade, zeigt aber keine einheitliche Krümmung, die auf eine parabelähnliche Gestalt der ganzen Einkommenskurve schließen ließe. Nur in den Fällen, bei denen auch die Pareto'sche Kurve schon etwas von der geraden Linie abweicht, tritt eine stärkere Krümmung als Grundtendenz hervor.

Eine solche konkave Form hat die Kurve der Einkommen regelmäßig bei den Besitzeinkommen und den Vermögen und bei den kleineren, öfters bei allen Arbeitseinkommen und bei anderen speziellen Einkommenszweigen; in einzelnen Fällen bei allen Einkommen überhaupt. Pareto hat in diesen Fällen die Konstante a eingeführt, so daß die Formel die Gestalt $\log N = \log H - \alpha \log (R + a)$. Bei der nicht kumulativen Kurve[2]) zeigt sich diese Krümmung in

[2]) Die Gleichung dieser ($\log Y$, $\log R$) Kurven ist $\log Y = \log \left(\frac{\alpha . H . R}{(R+a)^{\alpha+1}}\right) = \log \left(\frac{\alpha N . R}{R+a}\right)$, wobei α konstant ist; diejenige der empirischen Kurve mit wechselndem α ist auch hier $Y = \alpha N$.

noch verstärktem Maße, ebenso aber auch die Abweichung von den empirischen Kurven, welche, wie erwähnt, eine konkave Krümmung nur bei den kleinsten Einkommen, bei den mittleren aber eher eine konvexe Einbuchtung haben.

Diese Beobachtung führt zu der Vermutung, daß die Einkommen keine homogene Menge bilden, sondern aus verschiedenenartigen Gruppen zusammengesetzt sind, von denen eine, die der Besitzeinkommen, eine Anordnung zeigt, die dem Gauß'schen Gesetz zu entsprechen scheint.

In der Tat ist die Entwicklungsmöglichkeit für Besitz- und Arbeitseinkommen grundverschieden. Die Vermehrung des Besitzes ist unbegrenzt, andererseits kann er auch in jeder beliebigen Kleinheit existieren. Ein Minimum des Besitzes an Gebrauchsgegenständen ist in unserem Klima unentbehrlich,

aber auch dieses kann durch Schulden überwogen werden, deren Zinsen das sonstige Einkommen bis zu einer gewissen Grenze schmälern können.

Hiermit ist die Reihe der Hypothesen und Theorien nicht erschöpft, insbesondere möchte ich auf die Arbeiten Pearson's und Edgeworth's verweisen, die das Problem der Verteilung der Häufigkeit eines Kollektivgegenstandes in umfassender Weise behandeln. Aber schon die angeführten Beispiele dürften zeigen, wie fruchtbar die graphische Darstellungsform statistischer Größen sich erweist, auch wenn man ihre weitere Umgestaltung auf einfache Rechenoperationen, Logarithmieren und Bildung von Differenzialquotienten beschränkt.

Auch die im Folgenden aufgestellten Hypothesen sollen nur die Anregung sein, den Gegenstand auch unter dem Gesichtspunkt der Unterscheidung von Besitz- und Arbeitseinkommen zu betrachten. Eine erschöpfende induktive Prüfung wäre zwar im Gegensatz zu den besprochenen Hypothesen Pareto's und Furlan's möglich, sie würde aber außerordentlich schwierig und müßte trotzdem unvollkommen sein, weil die meisten Statistiken sich weder auf das steuerfreie Einkommen beschränken, noch die Besitz- und Arbeitseinkommen kombiniert darstellen. Es müßte also auf das Urmaterial zurückgegriffen werden, wozu naturgemäß nur diejenigen Stellen in der Lage sind, die über dieses und die nötigen Arbeitskräfte verfügen.

So mußte sich die induktive Bestätigung auf die Anführung bekannter, wenn auch nicht zahlenmäßig einheitlich festgestellter Tatsachen und auf die Übereinstimmung des Resultats mit der von Pareto und andern genügend erprobten gradlinigen Form des Einkommensdiagramms beschränken. Um nicht zu ermüden, ist aber stets die direkte Rede angewandt worden.

Ich glaube nicht, daß es möglich ist, das Problem mit einem Schlage zu lösen. Die Aufgabe ist, dem Verständnis näher zu bringen, wie der, nach Galton symmetrischen Kurve der Begabungen die, auch bei logarithmischer Behandlung noch höchst unsymmetrische, Kurve der Einkommen entsprechen kann. Eine solche Aufgabe kann am besten schrittweise gelöst werden. Mit Hilfe der Statistik können die einheitlichen Massen der Begabungen und Einkommen in Teile zerlegt werden, die einander näher entsprechen. Für diese Teile können dann die möglichen Formen der Wechselwirkung zwischen Einkommen und Begabung aufgezählt und kann festgestellt werden, in welcher Richtung die Kurve der Einkommen gegenüber der Kurve der Begabungen durch jeden einzelnen dieser Wirkungen verschoben werden muß. Da ein isolierendes Experiment in der Regel nicht möglich ist, kann nur das Gesamtergebnis aller Einwirkungen gemessen werden.

Es sind also die im Teil B aufgezählten Methoden der Zerlegung in Teilmassen, der gegenseitigen Mischung und Beeinflussung dieser, der Auslese und der Verschiedenheit des Erfolgs der abändernden Ursachen anzuwenden. Danach ergeben sich folgende Hypothesen:

1. Zu Grunde zu legen ist eine logarithmische Skala, die die relativen, nicht die absoluten Unterschiede als Maß der Abänderung erscheinen läßt. Dies ist die von Fechner gefundene allgemeine Regel, die nur da nicht berücksichtigt zu werden braucht, wo wegen der Geringfügigkeit der relativen Unterschiede die logarithmische und die nummerische Skala sich nur wenig unterschieden. Ihre Außerachtlassung bringt bei den Einkommen die ungeheure, graphisch nicht mehr darstellbare Asymetrie der Ammon'schen Kurve zustande.
2. Die Einkommen bilden keine einheitliche Masse, sondern zerfallen in Arbeits- und Besitzeinkommen. Zum Teil tritt ein leiser Knick schon bei der kumulativen Kurve, noch mehr bei der Häufigkeitskurve erkennbar

hervor (Fig. 11, Sachsen). Das Beispiel von Bologna (Fig. 12) zeigt aber auch, wie aus den deutlich gekrümmten Kurven der I. und II. cotegoria eine nahezu grade Linie aller Einkommen entstehen kann.

3. Die Arbeitseinkommen sind nicht den Fähigkeiten der Menschen proportional sondern ergeben sich aus der Häufigkeit der Fähigkeiten (Angebot) und dem Bedarf des Wirtschaftslebens (Nachfrage) nach den Preisbestimmungsgründen. Von diesen kommt
   a) auf Seite des Angebots vor allem in Frage, ob die betreffende Fähigkeit
      α) ein Naturprodukt ist und als solches Seltenheitswert besitzt oder ob sie
      β) durch Ausbildung erworben oder gesteigert werden kann und zwar zu welchen Kosten, und endlich
      γ) ob das Angebot durch Gesetz oder Sitte beschränkt ist.
   b) Auf Seite der Nachfrage ist wohl am wichtigsten der Umstand, welche Bedeutung die Qualität der Leistung für die Produktion hat. Bei Arbeiten, deren Qualität an sich nur in engen Grenzen schwanken kann, von deren Güte und Schnelligkeit nicht die Ausnutzung eines großen Kapitals und wertvoller Materialien oder gar der Erfolg der Leistung selbst abhängt, ist die Produktivität der Arbeit dem geleisteten Quantum nahezu proportional. Die Bedienung kostbarer Maschinen und Bearbeitung teurer Stoffe, die Leitung eines ganzen Betriebes, überhaupt jede verantwortliche und schwierige Leistung kann dagegen bei schlechter Ausführung großen Schaden bringen. Der Unternehmer wird daher nach einer vorzüglichen Arbeitskraft suchen, auch wenn solche infolge der Seltenheit der Begabung oder der Kosten der Vorbildung sehr teuer sind. Der Träger einer solchen Arbeitskraft kann daher Gehälter fordern, die den Lohn eines einfachen Arbeiters vielfach übertreffen, wenn er es nicht vorzieht, seine Fähigkeiten in einem eigenen Betriebe zu verwerten.

In der Statistik erscheinen nun als Arbeitseinkommen nicht alle Beträge, deren Erwerb auf Rechnung der Verwertung der eigenen Arbeitskraft zu setzen ist. Vielfach beschränkt sie sich auf Zusammenstellung der im engeren Sinne so genannten Lohneinkommen, in anderen Fällen umfaßt sie auch die Gehälter der Angestellten und Beamten und den Ertrag freier Berufe. Dagegen scheidet das gesamte Einkommen der selbständigen Unternehmer aus, auch wenn es fast ganz auf der Verwertung der eigenen Arbeitskraft beruht. So kommt es, daß die Tantième eines Mitglieds des Aufsichtsrats, der seine Stellung vor allem seinem Aktienbesitz verdankt und die Funktionen eines Unternehmers faktisch versieht, als Arbeitseinkommen registriert werden kann, der Verdienst eines nominell selbständigen, in Wirklichkeit von seinem Verleger abhängigen Hausindustriellen dagegen als Unternehmergewinn, und zwar in seinem ganzen Umfange.

Durch diese Ausscheidung der selbständig Arbeitenden wird also die Kurve der Arbeitseinkommen[1]) geschwächt. Es ist zu vermuten, daß diese Schwächung besonders bei den mittleren Einkommen eintritt, während die höchsten und niedrigsten auch in der Statistik also solche erscheinen. Hervorragende Fähigkeiten erfordern zu ihrer vollen Ausnutzung einen Betrieb von einem solchen Umfang, daß die Träger der höchsten Arbeitskräfte in der Regel nicht das Kapital besitzen, sich in einem solchen selbständig als Unternehmer zu betätigen. Zugleich läßt die Gesellschaftsform die Abhängigkeit, z. B. eines Direktors, wenig drückend erscheinen und ermöglicht die Beteiligung am Kapital und

damit am Gewinn und der Macht. Fähigkeiten, zu deren Ausnutzung ein mittlerer Betrieb genügt, können leichter zu Ersparnissen führen oder mit einem Vermögen zusammentreffen, das zur Gründung eines solchen Betriebes ausreicht. Bei den geringsten Fähigkeiten endlich ist die Unternehmerstellung pekuniär eher von Nachteil und fehlt die Möglichkeit von Ersparnissen in der Regel.

Die Ausscheidung der mit Unternehmergewinn verbundenen Arbeitserträge wird also vermutlich zur Folge haben, daß die mittleren Arbeitseinkommen weniger zahlreich sind, also sie an sich sein würden. Diese Vermutung wird durch die Kurve des Kantons Waadland bestätigt und ist nach Pareto in der Regel zutreffend.

Werden auch die höheren Arbeitseinkommen ohne Unternehmerstellung ausgeschieden, so bildet der verbleibende Rest wieder eine konkave Kurve; wir nähern uns damit wieder der nach dem Gauß'schen Gesetz und der Fechner'schen Regel zu erwartenden Form.

Auf die Gestaltung der konvexen wie der konkaven Kurve der Arbeitseinkommen wirken noch weitere Umstände ein. Erstere umfaßt die höheren Einkommen, bei denen der Erfolg erst nach einer längeren und kostspieligen Vorbereitung einzutreten pflegt. Während dieser Zeit arbeiten die betreffenden Personen nicht unproduktiv, sie erwerben Fähigkeiten, die vielleicht sehr rentabel sein werden; ihre Arbeit erzeugt aber noch kein statistisch nachweisbares Einkommen. Beginnt der äußere Erfolg, so wird er im Allgemeinen um so höher sein, je größer die Aufwendungen sind. Ist der Erfolg ungewiß, so wird in noch höherem Maße das Einkommen der Erwerbenden größer, ihre Zahl geringer, als es den Fähigkeiten der in diesem Beruf Arbeitenden entsprechen würde. In allen Fällen wird die konvexe Kurve in ihrem oberen Teil, d. h. für die größeren Arbeitseinkommen schmäler und gestreckter, als es den wirklichen Arbeitsleistungen entsprechen würde.

Auch die andere Seite der Arbeitseinkommen entspricht nicht genau den Fähigkeiten. Für die wirkliche Arbeitsleistung ist nicht nur die vermutlich symmetrisch abgestufte Kurve der Begabungen, sondern auch der Grad der Anspannung dieser Fähigkeiten entscheidend. In dieser Beziehung kann man zwei Typen von Arbeitern unterscheiden. Diejenigen, die Aussicht auf Verbesserung ihrer Lage haben und deren Bedürfnisse nicht durch das Herkommen beschränkt sind, arbeiten in der Regel um so intensiver, je größer der Erfolg ist. Sie steigen selbst zum Teil in höhere Stellungen auf oder versuchen, ihren Nachkommen ein Aufsteigen zu ermöglichen. Sie verteilen sich über die mittleren und oberen Stufen der Einkommenspyramide. An der unteren Grenze stehen aber diejenigen, deren Fähigkeiten zur Erhaltung ihres Lebens kaum ausreichen. Eine Möglichkeit, ihre Lage zu verbessern, besteht in der Regel nicht. Ihre Bedürfnisse mögen zwar einer unbegrenzten Steigerung fähig sein, können aber nur mit übermäßigen Anstrengungen befriedigt werden, oder sind überhaupt nur unerfüllbare Wünsche. Praktisch in Frage kommen nur die Bedürfnisse der Erhaltung des Individuums und der Familie. Diese steigern sich bei Nichtbefriedigung bis zur Unerträglichkeit. Daher wird ein solcher Arbeiter alles daran setzen, um das Notwendige zu erwerben, und dazu seine Kräfte aufs äußerste anstrengen, wenn sie gering sind. Ist seine Leistungsfähigkeit größer, so wird er sie vor allem dazu verwenden, die notwendige Leistung mit geringerer Anstrengung, unter Schonung seiner Gesundheit, auszuführen, wird nach kürzerer Arbeitszeit, besserer Nahrung, Kleidung, Wohnung, Erhöhung streben, ehe er daran denken kann, seine oder seiner Nachkommen Lage zu verbessern. So erklärt sich die z. B. von Marie Bernays beobachtete große Verschiedenheit

der Ermüdung, der Arbeitslust, also anscheinend auch der Arbeitsfähigkeit bei Arbeitern von wenig abweichendem Arbeitserfolg und Lohnhöhe.

Die beiliegenden Diagramme geben eine schematische Vorstellung, wie die Verschiedenheiten der Bedürfnisse und Fähigkeiten zu einer außerordentlich verschiedenen Anspannung der Arbeitskraft führen können, und wie im Endergebnis die Kurve der Arbeitsleistungen dadurch beeinflußt wird.

Die Notwendigkeit, das Existenzminimum zu erringen, würde zu einer noch viel gewaltsameren Anspannung der Arbeitskraft und bei Mißerfolg zur Vernichtung des Lebens führen, wenn nicht die Familie und die Gesellschaft die ungenügende Leistungsfähigkeit der Einzelnen ergänzen könnten. Nur so ist die Erhaltung des Menschen in der Kindheit, bei Krankheit und im Alter in der Regel gesichert. Durch die Wirtschaftseinheit der Familie wird die Tätigkeit des Menschen in der Kindheit, die im höheren Sinne ja außerordentlich produktiv ist, wird die Erhaltung der Menschheit selbst erst möglich gemacht.

In der Statistik erscheinen diese Fähigkeiten nicht als Einkommen, so hoch auch die ideellen und materiellen Werte sein mögen, die eine Mutter und Hausfrau schafft. Nur der viel geringfügigere Lohn eines Lehrlings oder jugendlichen Arbeiters, einer Tochter, die, um zum Haushalt beizutragen, in der Fabrik arbeitet, gelten als Einkommen, weil sie von außerhalb der Wirtschaftseinheit herstammen. Wollte man sich schätzungsweise einen Begriff der gesamten Produktion und Konsumation machen, einschließlich derjenigen Güter, die innerhalb der Familie erzeugt und verbraucht werden,[1]) so würde ihr Umfang weit über

[1]) Nach Kiaer beträgt der Wert der hauswirtschaftlichen Produktion Norwegens 136 000 000 Kr. bei ca 600 Mill. Kronen Gesamteinkommen und 2 Millionen Einwohnern.

das hinausgehen, was man als „Volkseinkommen" zu bezeichnen pflegt. Gerade die wertvollsten Güter entziehen sich der Statistik; nicht die Erhaltung der Gesundheit, sondern die Versuche ihrer Wiederherstellung, nicht das gesicherte Recht, sondern seine Verletzung bringt statistisch erkennbare Tatsachen hervor.

So ist es unmöglich, daß die Kurve der Begabungen in der Kurve der Einkommen ihr vollkommen entsprechendes Gegenstück finden sollte.

Aus dem bisherigen erklären sich auch die Abweichungen, die spezielle Lohnkurven zeigen. So läßt z. B. die amerikanische Lohnstatistik von 1902, die für eine Reihe von Gewerben Wochen und Stundenlöhne für die einzelnen Arbeitekategorien und das ganze Gewerbe gibt, durch ihre „cumulative percentage column" mit ziemlicher Übersichtlichkeit die Gestaltung der Lohnhäufigkeiten erkennen. Für die ganzen Gewerbe liegt das Maximum in der Regel bei den niedrigen Löhnen, weil die geringbezahlten Arbeitergruppen die zahlreichsten sind. Innerhalb der Gruppen liegt die größte Häufigkeit jedoch höher, oft über dem Zentralwert. Die graphische Darstellung dieser Statistik würde die Gesetzmäßigkeiten leichter erkennen lassen, ohne die Einzelheiten zu verwischen, wie es durch die Zusammenstellung der Quartilen und Mittelwerte geschehen mußte.

Alle diese Momente lassen es als begreiflich erscheinen, daß die Kurven der individuellen Arbeitserträge sich von denjenigen der Begabung unterscheiden und je nach der Abgrenzung des Beobachtungsfeldes untereinander abweichen.

Eine weitere Modifikation tritt bei denjenigen Statistiken ein, die sich auf die Veranlagung der Wirtschaftseinheiten zur Einkomensteuer stützen. Für die Wirtschaftseinheit gilt ferner das Existenzminimum in seiner ganzen Schärfe. Es ist allerdings verschieden, nach der Anzaxhl der Mitglieder der Familie oder sonstigen Gemeinschaft, es zieht aber eine Grenze, unterhalb der der Abgrund der Hilfsbedürftigkeit, der Abhängigkeit, Entbehrung und des Verbrechens beginnt. Mit allen Kräften strebt jeder, der es vermag, sich über diese Grenze zu

erheben, andernfalls trifft zwar nicht den Einzelnen das Los der Vernichtung, wohl aber die Familie Entartung und Erlöschen. Durch Sterblichkeit und ungenügende Entwicklung der Kinder, Krankheiten, Schwäche der Erwachsenen, entsteht wieder vermehrter Bedarf und geringeres Einkommen, bis die unterste Stufe des Elends erreicht ist und die soziale Auslese ihr Werk vollenden kann. So bildet die „Grenze der Brauchbarkeit" auf die Dauer eine wirksame Schranke, die über sich die Häufigkeit anschwellen läßt, unter sich aber die Herabsinkenden der Zerstörung preisgibt.

So erklärt sich für die Wirtschaftseinheiten der jähe Abschluß der Paretoschen Kurve.

Auf der anderen Seite erwächst aus den Überschüssen des Erwerbs über den Verbrauch der Besitz.

Würde man schematisch die Einkommen über einer gewissen Grenze absondern und annehmen, daß der Überschuß zur Kapitalsbildung verwendet werde, so würde die Kurve der Besitzeinkommen sich aus der Arbeitskurve konstruieren lassen; bei arithmetischer Skala brauchte man nur das obere Stück abzutrennen, bei logarithmischer Skala wird der Übergang von 0 bis 1 unendlich groß, es würde also eine nach unten allmählich sich verengende Kurve entstehen; der statistisch erfaßte Teil liegt aber so hoch, daß er von der Kurve der Einkommen wenig abweichen würde.

Wollte man annehmen, daß alle Klassen einen gleichen Prozentsatz ihres ganzen Einkommens zur Ersparung benutzten, so würde die Kurve der Vermögen ein genaues Ebenbild der Kurve der Einkommen werden.

In Wirklichkeit sind die Vermögen weder den ganzen noch den um das Existenzminimum gekürzten, den „freien" Arbeitseinkommen proportional; ihre Kurve ist konkav. Dies erklärt sich zunächst dadurch, daß nicht nur die in der Statistik als solche aufgeführten Arbeitseinkommen, sondern vor allem diejenigen, die in eigenen Unternehmungen erworben werden und als Gewerbe oder Bodenertrag registriert werden, zur Vermögensbildung beitragen. Diese ergänzen, wie schon ausgeführt, die Kurve der Arbeitseinkommen zu einer konkaven Linie. Fügt man die Unternehmergewinne hinzu, die besonders die mittleren Arbeitseinkommen relativ stark erhöhen, so wird die Krümmung der Linie noch höher.[1])

[1]) So in Bologna, vergl. beiliegende Figur 12.

Die so gebildeten Vermögen werden im Wirtschaftsleben verwendet und den Chancen des Gewinnes und Verlustes ausgesetzt . Diese abändernden Ursachen ihrer Größe sind in ihrer Mannigfaltigkeit unübersehbar. Ihre gegenseitige Kompensation müßte die Krümmung der Vermögenskurve noch verstärken. Entgegengesetzt wirkt der von Gini[2]) nachgewiesene Umstand, daß

[2]) Il diverso accrescimento delle classe sociali e la distrobuzione della riechezza.

dis Familien mit größerer Nachkommenschaft im Durchschnitt weniger wohlhabend sind, und so die Verschiedenheit der Einkommen noch gesteigert wird.

Unter Zusammenfassung der bisher dargestellten Hypothesen ergibt sich folgender Erklärungsversuch für die Regelmäßigkeit der Einkommensverteilung:

1. Die wirtschaftlichen Fähigkeiten der Menschen werden durch Vererbung mit geringen positiven oder negativen Abweichungen übertragen; dadurch entsteht nach dem Gesetz der Wahrscheinlichkeit eine symmetrische binomische Kurve und zwar mit logarithmischer Skala, da gleiche relative Unterschiede gleiche Bedeutung haben. (Fechner, K. M. L.)
2. Die Menschen alssen sich so in Gruppen einteilen, je nachdem sie durch ihre Fähigkeiten imstande sind

a) nicht einmal ihr Leben zu erhalten,
b) durch Arbeit ihr Leben zu fristen,
c) eine Familie zu ernähren,
d) eine angesehenere Stellung zu erringen,
e) höhere Bedürfnisse zu befriedigen.

3. Die Gruppe a und die Angehörigen der Gruppe b sind auf Wohltätigkeit angewiesen oder fallen dem Verbrechen anheim; sie verlieren ihre Arbeitsfähigkeit noch mehr durch Arbeitslosigkeit und Entbehrungen, und vermindern sich durch Überschuß der Todesfälle über die Geburten, soweit sie nicht durch herabsinkende Elemente ergänzt werden.

Die Gruppe c vermehrt sich am stärksten und bildet den Typus der folgenden Generationen.

Die Gruppe d ist die sozial aufsteigende Klasse. Sie strebt nach Macht, und zum Zweck und mit Hilfe derselben nach Besitz. Dieses Streben konkurriert mit dem Wunsche nach Nachkommenschaft und beschränkt ihn.

Die Gruppe e differenziert sich am stärksten, da die höheren Bedürfnisse wachsen, wenn die die aus ihnen entspringenden Wünsche befriedigt werden. Ein Teil, e α verfolgt das Streben nach Ehre, Macht und Reichtum um des Strebens willen, andere — e β — wenden sich religiösen, ästhetischen, wissenschaftlichen usw. Neigungen zu. Der Wunsch nach Nachkommenschaft wird zurückgedrängt, die Erwerbsfähigkeit der Teilgruppe e β gemindert und ihre Nachkommen sinken in die früheren Gruppen zurück.

So besteht nur bei einer Minderheit (d + e α) eine Vermehrung des Besitzes, die nur bei e α unbegrenzt ist.

4. Die Nachkommen aller Gruppen stehen unter dem Einfluß der Vererbung, d. h. die Eigenschaften der Vorfahren sind bei ihnen häufiger vertreten als im Gesamtdurchschnitt, die Abweichungen um so seltener, je größer sie sind. Aus dieser Tendenz allein würde folgen, daß jede Klasse einen Teil ihrer Angehörigen an die nächst höhere, einen andern an die nächstniedrige abgibt, und sich das binomische Diagramm verbreitert, indem die mittleren Klassen mehr abgeben, die äußeren mehr empfangen.

Da sich aber in den Nachkommen die Vererbung zweier Eltern kombiniert, ist die Differenzierung wieder eingeschränkt, und zwar umsomehr, je verschiedener die Eigenschaften der Eltern sind. Würden sich die Eltern gleichmäßig aus allen Gruppen der Gesellschaft verbinden, so würde in wenigen Generationen jede Differenzierung der ererbten Eigenschaften verschwinden; nur dadurch, daß die Verbindungen sozial, wirtschaftlich, persönlich ähnlicher Individuen verhältnismäßig leichter sich realisieren, als die Verbindungen verschiedener Individuen, kann sich eine erbliche Differenzierung erhalten.

Es kommt zur Bildung von Klassen und sozialen Kreisen, bei denen nicht nur durch die Kombination der Eigenschaften der Eltern sich ein Typus der Vererbung konzentriert, sondern auch die Lebensgewohnheiten und Anschauungen sich gegenseitig anpassen, und zwar um so stärker, je mehr Sitte und Herkommen für die Lebensführung der Einzelnen bestimmend sind.

Die so sich ergebenden Lebensbedingungen wirken wieder auf die Zahl und die Eigenschaften der Nachkommenschaft: Diejenigen der Gruppen a und b und zum Teil der Gruppe e degenerieren, diejenigen der Gruppe c passen sich überwiegend den Lebensbedingungen an und erwerben steigende Leistungsfähigkeit, die ihnen das Aufsteigen in die Klasse d ermöglicht.

5. Durch die Trennung der Gesellschaft in Klassen von verschiedener Lebenshaltung kann ein nominell gleiches Einkommen in der einen Klasse nicht einmal zur selbständigen Wirtschaftsführung einer einzelnen Person ausreichen, das in seiner anderen Klasse nicht nur Erhaltung der Familie und Mehrung des Besitzes an materiellem Kapital, Gebrauchsgütern und persönlichen Fähigkeiten, sondern auch Befriedigung höherer Bedürfnisse ermöglichen würde. Die Einteilung der Gruppen a—e durchkreuzt also die der sozialen Klassen; ein Individuumkann unter Umständen durch Herabsinken in eine als niedriger geltende soziale Klasse in eine höhere Gruppe der Bedürfnisbefriedigung aufsteigen. Eine Massenerscheinung dieser Art ist die Aufgabe wirtschaftlicher Selbständigkeit, um als Arbeiter und Angestellter im Großbetrieb eine materiell günstigere Stellung zu erhalten. Auch das Umgekehrte ist von Bedeutung: Materielle Entbehrungen und Verzicht auf Familiengründung, um sozial aufsteigen und höhere Bedürfnisse befriedigen zu können. Führende Persönlichkeiten auf allen Gebieten, Bahnbrecher neuer Ideen und Bewegungen sind diesen Weg gegangen und viele sind ihnen gefolgt. Ihre Zahl ergänzt sich immer aufs neue, aber zu einer Klassenbildung gelangen sie nicht, denn dem erfolgreichsten Teil gelingt es, in die höhere Klasse aufgenommen zu werden, die meisten bleiben in einer Lage, die ihnen die Gründung von Familien unmöglich macht.

6. Besitz- und Arbeitseinkommen stehen in einer elastischen Proportion zu einander.
   a) Wächst der bewegliche Besitz schneller als die Bevölkerung, und schneller als der Kapitalbedarf infolge verbesserter Technik, so sinkt der Zinsfuß. Dies hemmt das Wachsen des Einkommen aus beweglichem Kapital zugunsten des Arbeitseinkommens;
   b) die Zunahme der Bevölkerung und ihres Reichtums steigert die Grundrente;
   c) das Sinken des Zinsfußes findet seine Grenze durch sich selbst, indem es die Entwicklung der Technik befördert. Diese führt zu neuem Kapitalsbedarf, zur Verwertung von Erfindungen, welche
   d) neue Naturkräfte, neue Stoffe und neuen Boden verwendbar machen und so die Rente der alten schmälern.

So wird das Einkommen aus Besitz und dasjenige aus Arbeit durch die Bedingungen unserer Wirtschaftsordnung daran verhindert, durch Vermehrung der eigenen Produktiv-Kräfte anzuwachsen, ohne den anderen Einkommenszweigen einen Teil des Mehrertrages zukommen zu lassen. Damit erklärt es sich, daß in der Regel der relative Anteil der Arbeit und der des Besitzes am Ertrag von der Norm nicht allzuweit abweichen, und aus der für größere Gebiete nicht einseitig die Kurve aller Einkommen beeinflussen können.

Sie hat daher die Tendenz, sich oberhalb des dichtesten Wertes einer geraden Linie zu nähern, wie es die Erfahrung mit Hilfe der Diagramme Pareto's bestätigt hat.

Anhang zu C

Gleichungen nach Pareto, Gini, L. March usw.

Es empfiehlt sich, für die in Frage kommenden Größen einheitliche Bezeichnungen zu wählen, und mit den in den Originalwerken gebrauchten zusammenstellen. Es sei, im Wesentlichen entsprechend Gini:

	bei Pareto 1895	später	Gini
R die jeweilige Einkommenshöhe (= reddito)	X	X	R
N die Zahl der Censiten mit Einkommen von R und darüber (= numero)	—	N	N
A das Gesamteinkommen dieser Personen (= ammontare)	S ∞	R	A
H Konstanter Faktor in der Gleichung $N = \frac{H}{R^\alpha}$	-	A	H
K Konstanter Faktor in der Gleichung $N = \frac{A^\delta}{K}$	—	—	K
α Konstanter Exponient in der Gleichung $N = \frac{H}{R^\alpha}$	h	α	α
$\beta = (\alpha - 1)$	—	—	β
$\gamma = (\alpha + 1)$ „ „ „ „ $y = \frac{\alpha H}{R^{\alpha+1}}$	h	—	—
δ „ „ „ „ $N = \frac{A^\delta}{K}$	—	—	δ
a, b, andere (Krümmungs-) Konstanten	—	a,β	—
$y = \frac{-dN}{dR}$ Häufigkeit der Censiten	y	—	—
$z = \frac{-dN \cdot R}{dR} = y\,R.$			
$y\,dh = \frac{-dN}{dR}\,dR$ Zahl der Censiten zwischen R und (R + d R)	y dx	—	—
$z\,dR = \frac{-dA}{dR}$ Summe der Einkommen derselben			

Es lautet danach die Pareto'sche Gleichung in ihrer einfachsten Form, die für die kumulativen Reihe aller Einkommen annähernd zutrifft:

(1)
$$N = \frac{H}{R^\alpha}$$

Die Ableitung von N nach R gibt die von Pareto 1895 im Giornale degli Economisti aufgestellte Formel

(2)
$$y = \frac{-dN}{dR} = \frac{\alpha H}{R^{\alpha+1}}$$

Die Zahl der Censiten zwischen R und R + d R ist

$$y\,dR = \frac{-dN}{dR} \cdot dR = \frac{\alpha H\,dR}{R^{\alpha+1}}$$

Multipliziert man diese mit R, so erhält man die Summe ihrer Einkommen

$$R \cdot y \cdot dR = \frac{\alpha \cdot H \cdot R \cdot dR}{R^{\alpha+1}} = \frac{\alpha H \cdot dR}{R^\alpha}$$

Um diese Größe nimmt A ab, wenn R auf (R + d R) steigt; es ist also

$$\frac{-dA}{dR} \cdot dR = R \cdot y \cdot dR = z \cdot dR = \frac{\alpha H\,dR}{R^\alpha}$$

Integriert man z d R, so erhält man A für die Einkommen von R_0 bis R_1

$$A = \int_{R_1}^{R_0} z\,dR$$

Gilt die Pareto'sche Formel nach oben unbegrenzt, so ist

$$A = \int_\infty^R z\,dR = \int_\infty^R \frac{\alpha H\,dR}{R^\alpha} = \frac{\alpha}{\alpha-1}\int_\infty^R \frac{(\alpha-1) \cdot H\,dR}{R^{(\alpha-1)+1}} = \frac{\alpha}{\alpha-1} \cdot \frac{H}{R^{a-1}}$$

(3)
$$A = \frac{\alpha H}{(\alpha-1)\,R^{\alpha-1}}$$

(4)
$$z = \frac{\alpha H}{R^\alpha} = \alpha N$$

Aus (3) ergibt sich für R:

$$\frac{1}{R^{\alpha-1}} = \frac{(\alpha-1)\,A}{\alpha\,H}\,;\quad \frac{1}{R^\alpha} = \left(\frac{(\alpha-1)\,A}{\alpha\,H}\right)^{\frac{\alpha}{\alpha-1}}$$

Setzt man letzteren Wert in (1), so hat man:

$$N = H \cdot \left(\frac{(\alpha-1)\,A}{\alpha\,H}\right)^{\frac{\alpha}{\alpha-1}} = \left(\frac{1}{H}\right)^{\frac{1}{\alpha-1}} \cdot \left(\frac{\alpha-1}{\alpha} \cdot A\right)^{\frac{\alpha}{\alpha-1}}$$

Immer unter Voraussetzung genauer Geltung der Formeln müßte mit dieser Gleichung die Gini'sche

$$N = \frac{A^\delta}{K}$$

übereinstimmen; danach wäre

$$\delta = \frac{\alpha}{\alpha-1}$$

$$K = H^{\frac{1}{\alpha-1}} \cdot \left(\frac{\alpha}{\alpha-1}\right)^{\frac{\alpha}{\alpha-1}} = \delta^\delta \cdot H^{\delta-1}$$

Aus (3) ergäbe sich ferner

$$A = \delta\,N\,R$$

Im beiderseitig logarathmischen Diagramm muß man die Häufigkeit auf logarithmische Intervalle beziehen, wobei die Ableitung für natürliche Logarithmen bequemer ist. Es sei $Y = \frac{-dN}{d\,l\,R}$; $Z = \frac{-dR}{d\,l\,R} \cdot R$. Nun ist $N = H \cdot e^{-\alpha\,l\,R}$,

daher $\frac{dN}{d\,l\,R} = -\alpha N$; $\quad Y = \alpha N$

$$Z = \alpha \cdot N \cdot R$$

Durch Integration erhält man wieder

$$A = \int Z\,d\,l\,R = \frac{\alpha H}{\alpha-1}\int \frac{(\alpha-1}{R^{\alpha-1}}\,d\,l\,R = \frac{\alpha}{a-1} \cdot \frac{H}{R^{a-1}} = \delta \cdot N \cdot R.$$

Umgekehrt ist

$$Z = \frac{-dA}{d\,l\,R} = \alpha N R = (\alpha-1)\,\frac{\alpha}{\alpha-1}\,N R = (\alpha-1)\,A = \beta A$$

Da die Gleichungen

$$Y = \alpha\,N;\quad Z = (\alpha-1)\,A$$

für jedes noch so kleine Stück einer geraden Linie gelten, und man die empirischen Kurven aus solchen gradlinigen Stücken zusammengesetzt denken kann, ist die Kurve der (log Y, log R) aus der Kurve (log N, log R) leicht zu erhalten, indem man zu log N jeweils log α addiert. α ist $\frac{-d \log N}{d \log R}$ und aus dem Neigungswinkel der Figur der Kurve zu ersehen. Ebenso ist $(\alpha-1) = \beta = \frac{-d \log A}{d \log R}$ und daher die Kurve (log Z, log R) aus (log A, log R) zu konstruieren.

Die so gewonnene Kurve müßte mit derjenigen der (log (R . Y), log R) zusammenfallen, wenn die Einkommen der Censiten auch innerhalb der Steuerstufen sich nach der Pareto'schen Formel verteilen würden. Aus der Abweichung ergibt sich, wie weit die Angaben im oberen Teil der Steuerstufen bevorzugt worden sind.

Fügt man in die Pareto'sche Formel die Konstante a ein, so lautet sie

$$N = \frac{H}{(R+a)^{\alpha}}$$

Die Ableitungen sind dann

$$y = \frac{-dN}{dR} = \frac{H}{(R+a)^{\alpha+1}} = \frac{N}{R+a}$$

$$Y = \frac{-dN}{d\,l\,R} = \frac{\alpha H}{(R+a)^{\alpha+1}} = \frac{\alpha N . R}{R+a}$$

$$z = \frac{\alpha N R}{R+a}$$

$$Z = \frac{\alpha \cdot N \cdot R^2}{R+a}$$

Die Lucien March'sche Formel lautet in derselben Schreibweise

$$y = \frac{A R^m}{a^R}$$

Für die Intervalle der natürlichen Logarithmen ist anstatt $y = \frac{-dN}{dR}$ zu setzen $Y = \frac{-dN}{d\,l\,R} = y . \frac{dR}{d\,l\,R} = y . R$

$$Y = \frac{A R^m . R}{a^R} = \frac{A . R^{m+1}}{a^R}$$

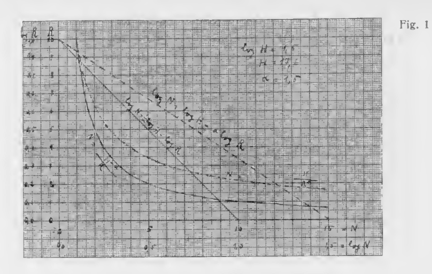


Fig. 1

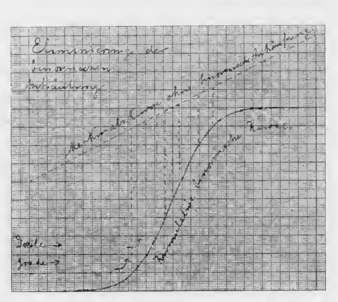


Fig 2

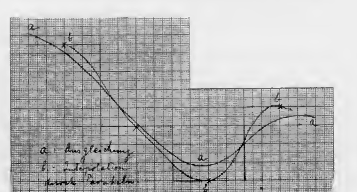


Fig. 3

Die so gewonnene Kurve müßte mit derjenigen der (log (R . Y), log R) zusammenfallen, wenn die Einkommen der Censiten auch innerhalb der Steuerstufen sich nach der Pareto'schen Formel verteilen würden. Aus der Abweichung ergibt sich, wie weit die Angaben im oberen Teil der Steuerstufen bevorzugt worden sind.

Fügt man in die Pareto'sche Formel die Konstante a ein, so lautet sie

$$N = \frac{H}{(R+a)^{\alpha}}$$

Die Ableitungen sind dann

$$y = \frac{-dN}{dR} = \frac{\alpha H}{(R+a)^{\alpha+1}} = \frac{\alpha N}{R+a}$$

$$Y = \frac{-dN}{d\,l\,R} = \frac{\alpha H R}{(R+a)^{\alpha+1}} = \frac{\alpha N . R}{R+a}$$

$$z = \frac{\alpha N R}{R+a}$$

$$Z = \frac{\alpha \cdot N \cdot R^2}{R+a}$$

Die Lucien March'sche Formel lautet in derselben Schreibweise

$$y = \frac{A R^m}{a^R}$$

Für die Intervalle der natürlichen Logarithmen ist anstatt $y = \frac{-dN}{dR}$ zu setzen $Y = \frac{-dN}{d\,l\,R} = y \cdot \frac{dR}{d\,l\,R} = y \cdot R$

$$Y = \frac{A R^m . R}{a^R} = \frac{A . R^{m+1}}{a^R}$$

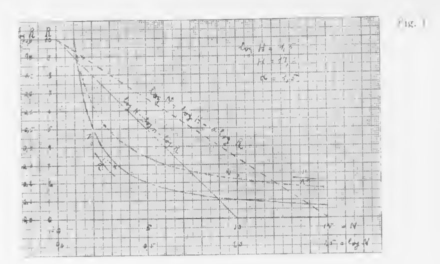


Fig. 1

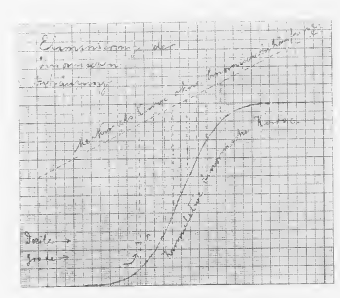


Fig. 2

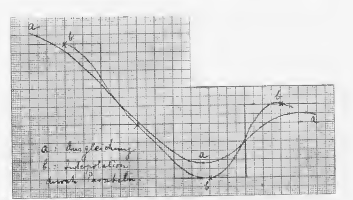


Fig. 3



Fig 4

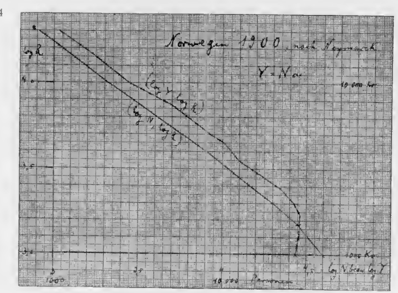


Fig. 5

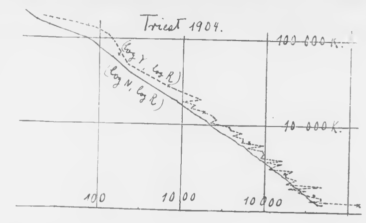


Fig 6

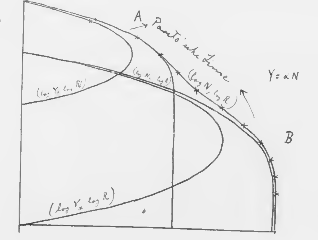


Fig. 9

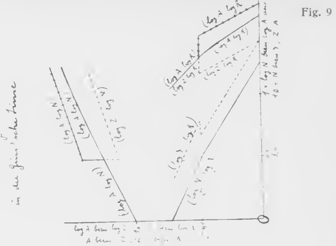


Fig. 7

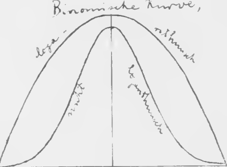


Fig. 8

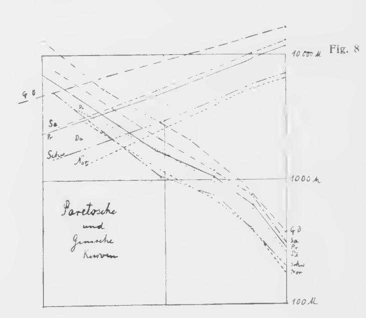


Fig. 4

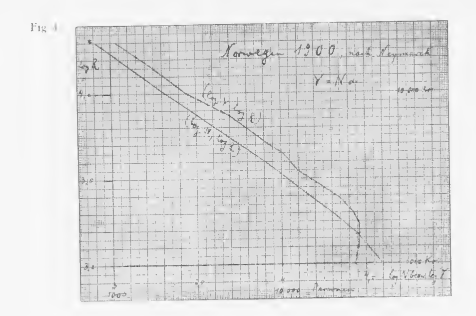


Fig. 5

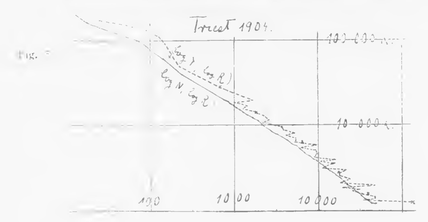


Fig. 6

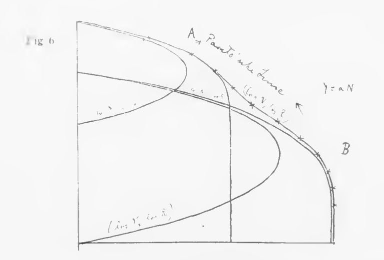
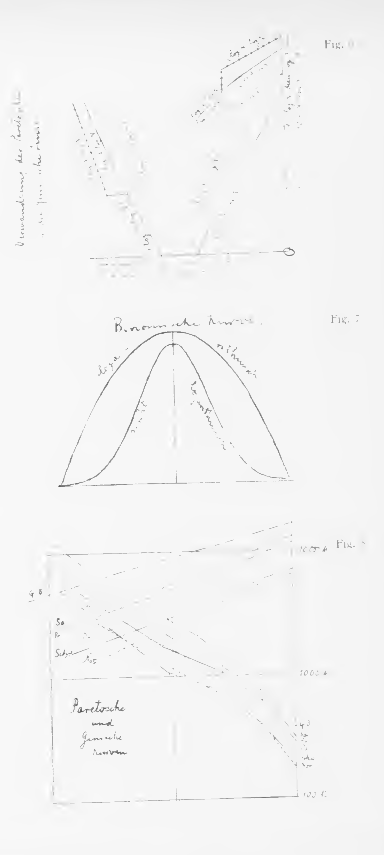


Fig. 6a

Fig. 7

Fig. 8



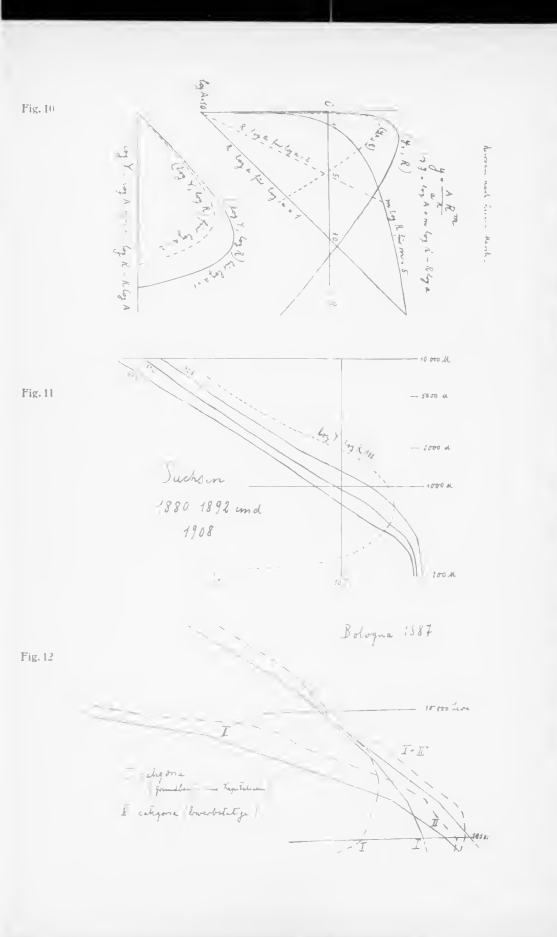


Fig. 10

Fig. 11

Fig. 12

Gedruckt bei A. Heindl & Co.
München S-W 2.

END OF TITLE